Hotmap: Looking at Geographic Attention

Danyel Fisher

Abstract—Understanding how people use online maps allows data acquisition teams to concentrate their efforts on the portions of the map that are most seen by users. Online maps represent vast databases, and so it is insufficient to simply look at a list of the most-accessed URLs. Hotmap takes advantage of the design of a mapping system's imagery pyramid to superpose a heatmap of the log files over the original maps. Users' behavior within the system can be observed and interpreted. This paper discusses the imagery acquisition task that motivated Hotmap, and presents several examples of information that Hotmap makes visible. We discuss the design choices behind Hotmap, including logarithmic color schemes; low-saturation background images; and tuning images to explore both infrequently-viewed and frequently-viewed spaces.

Index Terms—Geographical visualization, GIS, heatmap, server log analysis, online mapping systems, social navigation.

1 Introduction

Online interactive mapping systems are growing in popularity: Microsoft, Google, and Yahoo all offer interactive maps, with varying features and qualities of aerial photography. For the maintainers and designers of such systems, the behavior of users can provide critical cues on how both to improve their offerings and understand how users are now interacting with the system.

The tools needed to track user data for these systems are different from the traditional tools of log file analysis. Maps have the advantage of being tied to common-sense geography: a world map. In contrast, many log-file analysis tools must create an abstract space in which to situate a node-link diagram representing the website [9][15]. Mapping systems lack hyperlinks: while other online systems logically lead users between pages, online maps can be scrolled and navigated without clicking links. Applying a log-file analysis tool to these systems fails to represent the relevant dimensions of a mapping system: reconstructing a path between pages by examining referrer logs, for example, cannot apply. A simple reporting of the most-requested page requires substantial interpretation: presenting the filenames for individual map images is far from intuitive. Instead, the analysts' understanding of geography can be harnessed to build meaningful log-file visualizations by showing the logs as a map.

1.1 Harnessing Users' Sense of the World

A map of users' interaction with a space provides information about how they understand and interact with the world. What places do they find most significant? Where do they look at maps of roads, and where do they look at aerial imagery? How do they move through the space of the map?

There is a long tradition of research studying how people perceive and interact with the physical spaces around them. Kevin Lynch [11] interviewed people within three cities, asking them to sketch parts of the city that they were familiar with. He found that Boston was highly "imageable"—people clearly knew where they were and how neighborhoods were oriented—while other cities, such as Los Angeles, were far harder to understand. (Even within Boston, some areas were better defined than others). William H Whyte [23] photographed patterns of human interaction from above, looking at where people chose to sit or walk, in order to understand how people socially understood and constructed the city. Both of these projects generated social images of physical places. The server logs analogously provide an image of how physical spaces are understood by the users of online mapping systems.

This paper presents *Hotmap*, which visualizes the use of a Microsoft's Live Search Maps, one such online mapping system. It presents the number of downloads of imagery to the user, placed accurately on a map. Hotmap uses a heatmap [19] to represent aggregate activity within the system. It is implemented as a mash-up over an online mapping system, and so provides the same interactivity that the original map does, and provides the user with a view of the current maps and imagery.

The primary contribution of this work is this coloring of an interactive map based on users' views of geographic areas, and a discussion of the way it is used. We consider this visualization as showing users' attention to the map.

2 Task Overview

We begin by discussing imagery acquisition, the task that drove the design process for Hotmap. While there are secondary tasks that have emerged from this project, many of the design decisions were made in conference with teams interested specifically in acquiring up-to-date imagery for an online mapping system. In this paper, the term "imagery" refers to the worldwide aerial and satellite photography that forms the core of these mapping systems.

Keeping up-to-date imagery is clearly important. While TerraServer—one of the first publically-available global mapping systems [1]—was able to collate archival USGS imagery, contemporary mapping systems require up-to-date imagery to stay competitive. Users want to both see pictures of and get directions to their freshly-built house and neighborhood; sudden attention can come from news events reshaping geography. Both the 9/11 attacks and Hurricane Katrina, for example, changed the ground where they happened; Microsoft garnered a degree of negative press by having stale imagery of Apple's headquarters [17].

Important as it is, imagery acquisition is weighted with a variety of design decisions and trade-offs. Imagery is aggregated from governments, from third-party vendors selling both satellite and aerial photography, and from commissioned flyovers; navigational information is aggregated from several map vendors. Each of these comes at a price. Some photographs can only be obtained during particular seasons in order to avoid clouds, snow, or the long shadows of steeply-angled sun. Some flyovers and imagery are limited by political or security issues. Nor is it a priority to keep every point recent: in some places, the world does not change

• *Danyel Fisher is with the VIBE Lab at Microsoft Research, E-Mail: danyelf@microsoft.com.*

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 27 October 2007.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

Figure 1. Hotmap showing southwestern Europe. A red mark shows that a tile was downloaded at least one time. Interest follows major cities, but also note the strong interest in coastlines. Imagery depicted is at level 13 (approximately 20 meters per pixel).

rapidly. Out of date imagery may be acceptable for a body of water or a forest.

Choosing the right photographs to take means predicting where users will want to look. A series of questions need to be asked: are they more likely to look to tourism-oriented sites, or their town's streets? Does usage follow population, or natural features? Are intermediate levels of imagery used mainly for way-finding in the maps, or do users linger on broader views? In addition, photography must be timed for upcoming demand: an Olympics may call for fresh imagery to ensure that the stadiums are well-represented. It is a high-stakes process: millions of dollars may be spent annually in acquiring imagery.

This is far from a mechanical task, and requires integrating a variety of data sources. Hotmap provides information to this process that can be used to help understand current user behavior and interaction, and so can act as a guide for imagery forecasting.

3 Related Literature and Techniques

Hotmap's geographic visualization of system logs draws on two simultaneous approaches. With the rise of the web, a variety of technologies have emerged to analyze and handle log files. In the last several years, as online mapping software has become broadly available, increasing numbers of systems have released APIs to allow user-generated data to be added to graphs.

3.1 Log Analysis

Traditional use analyses for online systems examine user behavior from a client, server, or proxy point of view. Many features are available commercially and have moved out of research labs. Each shares advantages and disadvantages: client-based and proxy-based systems (such as WebQuilt [9]) can be more difficult to deploy, although the instrumentation they yield can provide detailed information on the activities that the user is doing. Server-based retroactive log analyses can give detailed information about when a system was used and by whom, but can be confounded by intermediate proxies, firewalls, and aggregators that share IP addresses. A broad overview of such techniques is presented in [18]; retroactive log-file analysis is discussed broadly in [10].

In general, these techniques focus on tracking how users *moved through* the system: they emphasize a notion that users follow links from page to page, and attempt to track those transitions. The notion of "information scent" [2] extends this—it suggests that users looking at one web page might figure out where to look next based on information on the page. Systems like H3 [15] and TimeTube [3] can provide detailed graph- or tree-based visualizations of interaction with a website.

A different model comes from tracking how users looked at different portions of a web site, and displaying that information back on the site itself. Several projects [5][22] have overlaid recent use of a web server on to the web pages, separating use into pages that have been "very recently" and "not recently" read.

3.2 Geographic Visualization

Geographic visualization is a mainstay of information visualization [19]. The last several years have seen an explosion of data-driven mapping as free programming layers have become common. Annotating maps based on dynamically-collected internet-based data is a newer approach. Many of these have taken the form of "mash-ups" linking maps to external sources of data [7][16]. Map mash-ups come in two forms: discrete forms, such as pushpins and poly-lines, in which specific latitude/longitude data is specified and plotted; and overlaid raster images [6]. These have distinct uses: the discrete forms are powerful ways of highlighting a fixed number of points or outlining regions. In contrast, raster images allow the user to assign a value to every pixel. Most recent mash-ups have focused on discrete source data, such as home sales listings and public transit schedules.

Raster images, however, allow us to draw heatmaps based on spatial data. A heatmap assigns a color based on a value to each point on a map. Heatmaps are a familiar visualization for maps when data is well-defined for a dense set of samples over space. Heatmaps have also been used to portray social data, although this requires a process of interpolation or extrapolation, as most much information is keyed to survey areas: counties and voting districts [12][20].

Location-based and map-based systems, which can store latitude/longitude coordinates, have enabled a new type of heatmap visualizations. One recent project demonstrated a map showing the number of Flickr images geo-coded to that location [8]. Photogenic "hot spots" had many such images, while cooler spots had fewer. Mehler *et al* presented heatmaps of news stories' influence and distribution [13].

One interesting advantage of presenting this information as a mash-up over commercial systems is that they can be distributed easily on the internet for other users to interact with—indeed, they can even be mashed up themselves, superposing other sources of data over them.

4 How Users Use Hotmap

Microsoft's Live Search Maps is structured as a dynamic web site; as users interact with the map, the system downloads images at the margins of their current screen so they can scroll through the map. Individual components of this map are uniformly sized *tiles*, which are downloaded to the browser from tile servers; those servers generate logs. Hotmap visualizes these log files.

Hotmap is constructed as a mash-up over the original mapping system; users experience Hotmap as an interactive map. The colored map (as in Figures 1 and 2) shows the number of times that a tile at that particular location was requested: brighter points have had far more requests then duller points. The tiles represented on the map are of higher resolution than the map itself; we refer to the difference between them as an "offset." This offset is user-controlled.

This section presents several of the results from Hotmap, in two forms. First, we present narratives of how Hotmap is used based on conversations with its target users: analysts involved in imagery acquisition. Second, we present a series of novel insights that were found using Hotmap; those insights are illustrated in Figures 4 and 5.

4.1 Narratives of Hotmap Use

Hotmap has been deployed internally for slightly over a year, although it has been steadily evolving in conversation with several imagery teams. In order to better understand how tasks are done, three members of mapping teams described how they use Hotmap: "Walter[1]", who makes global decisions about purchasing imagery, "Nicholas," who is responsible for coordinating city-scale imagery acquisition; and "Jack," who is supervising the recent deployment of a mobile edition of the interactive map.

Walter is in charge of designing the plans for long-term imagery acquisition within Europe. Walter sets general goals for areas, deciding which cities and what areas to photograph, and so evaluates use throughout the world. "Internationally," he said, "it helps us to identify what we call wonders of the world, which we want to cover with high res[olution] satellite images. We can now make sure we are identifying the right places and also fine tune the actual coverage areas in those places." Walter pointed to Hotmap's view of southern Europe (Figure 1), and noted that users had a seeming strong interest in looking at coastlines, including areas where the coast is sparsely populated. A new goal, then, is to "make sure we cover not only the cities but all coastlines in decent resolution." While good quality certainly drives viewership, Hotmap allows Walter to compare places with similar coverage. Walter uses Hotmap both to understand how users see both broad worldwide imagery and also how they see individual cities.

Nicholas, who is responsible for acquiring city-scale imagery, uses Hotmap differently. In particular, Nicholas is given a prioritized list of cities, and decides which of those cities should be photographed and in what order. Within each city, he is responsible for choosing which areas are to be covered within a strict size budget. Hotmap gives him a cue for where to look when picking the most important areas to cover. For example, he referred to the usage patterns over Boston, Massachusetts (Figure 2). Nicholas pointed to the fact that in addition to the general bright spot over the center of downtown, there are bright spots around several other important areas: (**a**) Harvard University; (**b**) Fenway Park, a sports stadium; and (**c**) the Bunker Hill monument, an historical site. Nicholas interprets the number of hits—and thus the brightness—as "popularity," and so sees these as critical places to acquire further imagery. He also tries to learn users' preferences in well-covered cities so he can generalize to upcoming cities. Nicholas summarized his discoveries as "make sure to get the universities, stadiums, and airports." Nicholas then supplements this information with other online and offline sources.

Jack is a member of a team that has recently released a mobile

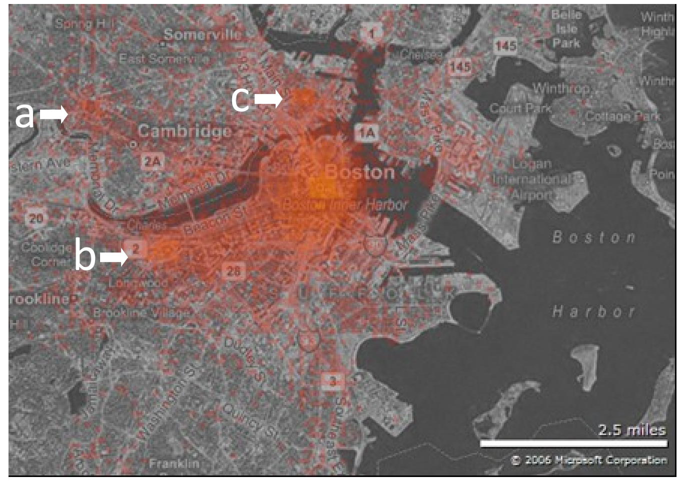

Figure 2. Hotmap showing the Boston area. Brighter spots depict tiles that have been downloaded more times. Note bright spots at (**a**) academic Harvard University, (**b**) athletic Fenway Park, and (**c**) the historic Bunker Hill Monument, in addition to downtown. Imagery depicted is at level 18 (60 cm per pixel).

application based on the mapping system. Jack's team is evaluating the use of their mobile tool, and would like to know where the tool is being adopted. Is it mainly being used to look at American roads, or has it been picked up more broadly around the world? This would help Jack's team decide on whether to concentrate on building internationalized versions of the tool rapidly. Jack's team provided usage logs from the early deployment of their system to Hotmap. As Figure 3 shows, Hotmap reflects experimentation with their application throughout the northern hemisphere—but much more in the English-speaking world. Internationalization of the interface might indeed help spur adoption.

4.2 Insights from Hotmap

Several other interesting phenomena help illustrate how Hotmap can be used. This section presents some insights from Hotmap, found both by teams using the tool and by researchers exploring the data. All the data samples in this section are based on a samplings of tile logs during 2006 and early 2007. Sampling methodologies varied during different stages of deployment, and so numbers of tile hits are relative to each other, and do not directly reflect traffic levels.

Conversations with internal researchers have suggested that users tend to start with several initial perspectives. The first, jokingly

Figure 3. Hotmap showing mobile requests. Imagery depicted is at level 17 (1.2 m per pixel).
Inset: one tile of level 17 imagery over New York City (reduced 50% for publication).

---

[1] All names are anonymized.

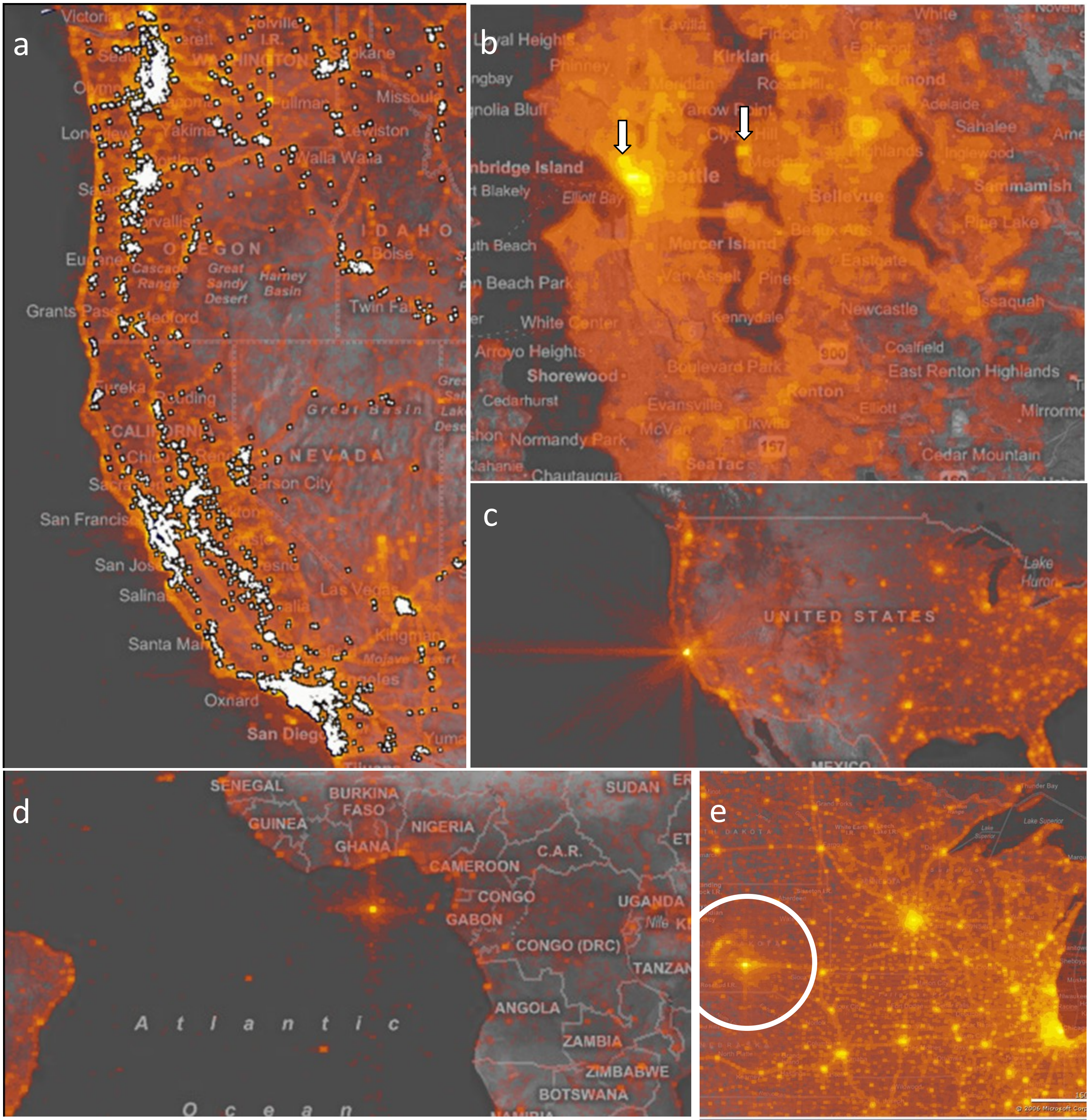

Figure 4. Insights from Hotmap. Top-left to lower right: (**a**) Hotmap representing the Western US (level *4*; *offset 7*) overlaid by a visualization of population density from the 2000 US Census [20] (white outlined areas). (**b**) The greater Seattle area (*level 10, offset 7*). Note relative bright spots at the Space Needle and Bill Gates' house (arrows). (**c**) The San Francisco star: a promotion drives users in straight lines (*level 3, offset 9)*. (**e**) The bright spot at latitude/longitude (0,0) *(level 3, offset 9)*. (**f**) The bright spot over South Dakota represents the default center of the US.

referred to as "I can see my house from here," involves a user looking up their own address and other familiar landmarks: a variant of Wattenberg's observations about social visualization [21]. The second links to tourism: users look up places they either have been, or wish to go. It is only after they are calibrated that users are prepared to start searching for restaurants, driving directions, and the like. Thus, we might expect Hotmap to generally reflect population, heavily weighted toward tourist locations.

4.2.1 The Relationship between Popularity and Population

In order to understand how Hotmap popularity compares to population, we superposed Hotmap with a gridded population view. Figure 4a compares Hotmap to the US Census' "Census 2000" [20] map, which represents every center of 7500 people in the United States as a single dot. Note that while tile popularity tends to follow population, users find some areas more interesting than population

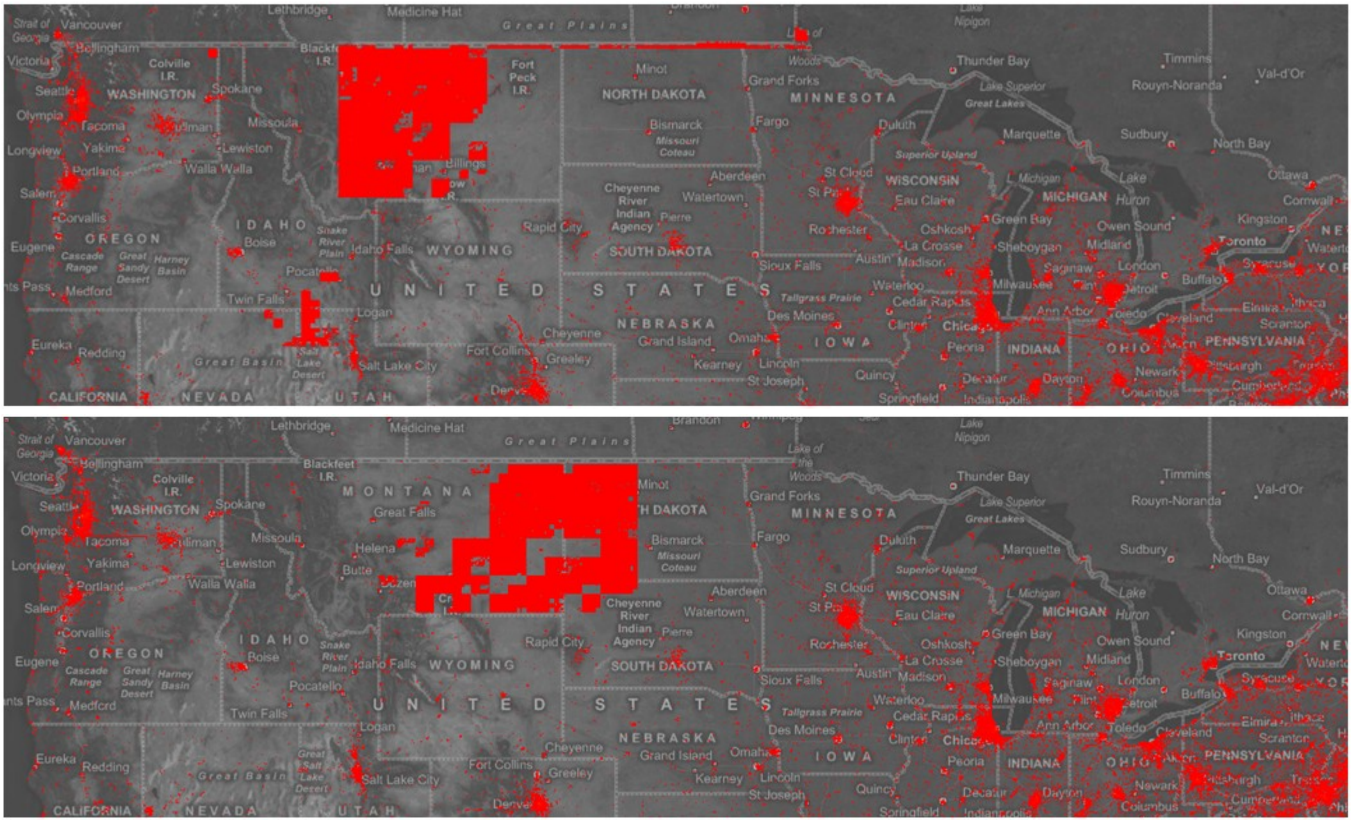

Figure 5. A scrape in progress. Two subsequent days of tile requests showing tiles that have received one or more hits across the northern US at zoom level 18, as seen from level 4. For scale, note that Seattle on the left, Philadelphia is in the bottom right corner.

would account for. In particular, Hotmap users tend to follow borders: roads (visible in figure 4e), shorelines, and even the bounds of available imagery. The census, shown as white outlines, comparatively shows the population distribution: Hotmap popularity generally follows population except in those places where some other factor drives additional views.

4.2.2 Tourism: A City with Prominent Points

It is valuable to interpret the most prominent points any given area, knowing that users look at them disproportionately to population. Note that the color scale is logarithmic; the users' focus is largely centered on a small number of locations. In Figure 4b, a view of the Seattle region, users find particular interest in a prominent landmark, the Space Needle (left), and in Bill Gates' house (right arrow). The latter came about because of a website link that pointed directly a view showing that address; the link was propagated and followed broadly.

4.2.3 Monitoring Promotions and Unexpected Effects

Shortly after updating Hotmap with a new dataset based on results from January, 2007, we were startled to see a "star" around San Francisco on Level 15 (Figure 4c). This looked like no other pattern that had been seen in this dataset. While Hotmap itself could not explain what had happened, one of the analysts was able to identify that a Flash banner ad had been promoting Virtual Earth's view of San Francisco during that time period. The banner ad contained a window showing the map, which followed the users' mouse: as the mouse moved on screen, the map would pan smoothly to keep up. The pattern in Figure 4c suggests the effects of users leaving their mice on the edge of the ad and simply allowing the system to scroll for a time: the brightness of the line suggests that this was common. This gives useful insight into how the advertising campaign was used—and suggests that alternate designs for the next one that might discourage infinite scrolling, and would stay within the bounds of imagery.

4.2.4 Finding Bugs and Usability Errors

There are some places that should not be popular; any users looking there are likely to be lost. Seeing many users there suggests a usability flaw or software bug. For example, the Atlantic Ocean somewhat south-west of the Ivory Coast has poor imagery and no population, but has a surprising bright spot (Figure 4d). This can be explained by noting that this point is on both the meridian and on the equator: it is (0,0) in the latitude/longitude coordinate system. A buggy application might accidentally send users there. In fact, an early bug in a location finding feature sometimes did fail by sending users to (0,0). While that bug has since been fixed, some mash-ups and user-created code may still lead users to this point. The four-pointed-star-shaped pattern is characteristic of "lost users" scrolling to try to find their locations.

A similar pattern may be seen over South Dakota (Figure 4e). For American users, the default map view centers over the United States. Should a user zoom before they pan, they find themselves zooming in over the middle of South Dakota. Confused users then pan the map in cardinal directions, trying to orient themselves.

4.2.5 System Security

The last section examined tiles that had unexpectedly high numbers of hits. Similarly, some patterns of tile downloading should also be unlikely to occur. Server analyses showed unusual activity levels of high zoom levels, and we wanted to understand what had caused it. In one mode, Hotmap can display what sections of the map *any* users have looked. One sample (Figure 5) showed that most users largely looked at city imagery for level 17. This is unsurprising: imagery is not available for most of the countryside at that scale. None the less, some user or users had downloaded vast areas of land systematically over a series of days (and, presumably, had been disappointed by the poor coverage). Lit up red, this cued the team to dig further for the source of this systematic scrape of the tile servers.

Figure 6. The tile pyramid. Tile 032 (highlighted) is at location x=2, y=3, level=3, and is one-fourth the size of tile 0 in each dimension.

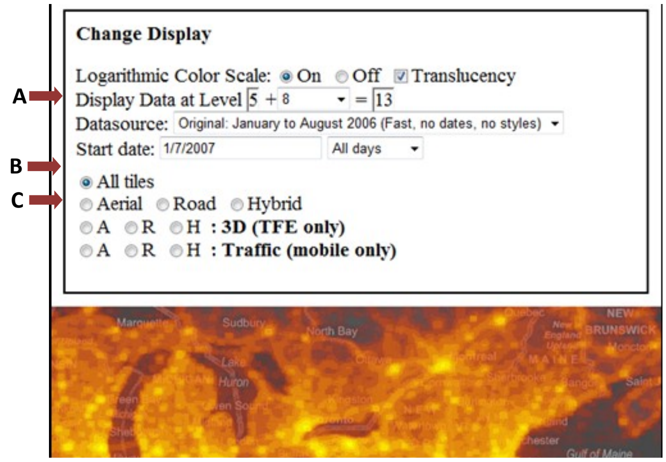

Figure 7. The Hotmap control panel.. **A** selects between imagery levels; **B** selects a date range**,** and **C** selects a data style.

5 CONSTRUCTING HOTMAP

Hotmap is constructed as an AJAX mash-up over Microsoft's Virtual Earth API. The Virtual Earth API includes an ability to superpose translucent tiles from a separate server. Constructing the system as a mash-up has several substantial advantages: as a web page, it can be deployed easily; other users can add further data sources over it; and its use can be measured with a server log. Hotmap consists of a page server, which serves the page around the embedded map; a tile server, which is responsible for returning tile images; and is backed by a database, which stores processed server logs.

The input to Hotmap is a sample of Virtual Earth tile server logs. These are trimmed to two core values: a date and a *tile identifier*. The tile identifier specifies both the location of the tile in space (x, y, and zoom coordinates) and the imagery style of the tile, such as "aerial" or "roads". For performance considerations, the date field is rounded to the nearest day. All tile requests with the same identifier on a given day are aggregated and stored in the database.

5.1 Data Confounds

There are several confounds in this data, based on browser, user, and cache behavior. The browser requests a screenful of tiles at a time, as well as a pre-cache border; the user will never see some of these. This means that a users' gaze is represented by at least a dozen adjacent hits, where the user was looking cannot be determined precisely. There are a series of other caches between the browser and server, and so all values that Hotmap presents are strict under-estimates. By definition, however, caching is likely to affect heavily-downloaded tiles more than unpopular ones; as such, the true popularity is spikier than the Hotmap view. While it is possible that Hotmap is missing some activity (perhaps systems that download and view tiles offline), the view does give us a good sense of the online behavior. Indeed, client caches mean that Hotmap is not distorted by some user behavior (such as a user repeatedly looking at the same point).

5.2 Hotmap itself generates a number of tile requests; should this number become significant, then it could generate a feedback cycle. This can be compensated for by examining the referrer log for the tile requests; tiles requested by the Hotmap client could be discounted.The Tile Pyramid

The Virtual Earth tile pyramid, described in [14] and illustrated in Figure 6, stores a zoomable image as static tiles, 256 pixels on a side. These tiles are stored hierarchically as a quadtree, with numbered levels from one (78 km per pixel at the equator) to 19 (30 cm per pixel). At the highest level, the entire image—in this case, much of the Earth's surface, excluding the poles—is stored as four tiles. Figure 6 illustrates this schema.

Hotmap replicates the tile pyramid representation of Virtual Earth for its own data store. This has the advantage of making it easy to render precisely: for level 8 imagery, in which a single pixel corresponds to one tile, the color of a pixel is exactly its degree of popularity. Storing information as latitude/longitude coordinates would require re-projection of the data to the underlying map.

The user interface of Hotmap presents a map and a selector for the level of detail (**A** in Figure 7): the user selects an "offset" relative to the current level of imagery. Hotmap computes the level of data to display by summing the offset with the current map view. For offsets up to eight, Hotmap checks the database for all records at the selected level of detail below the current tile, and colors them appropriately. After eight levels of offset, however, each record would be smaller than one pixel. For those, instead, Hotmap shows the single greatest value of the records that correspond to that pixel.

Figure 8 shows the same tile—tile 0, the top-left—overlaid with data at four offsets from the base tile, from four to eleven.

Because levels can have very different numbers of hits, and different semantic meaning, the visualization does not collapse levels together; a user can examine only one level of detail at a time. This distinguishes street-level imagery (level 18 and 19) from in-city

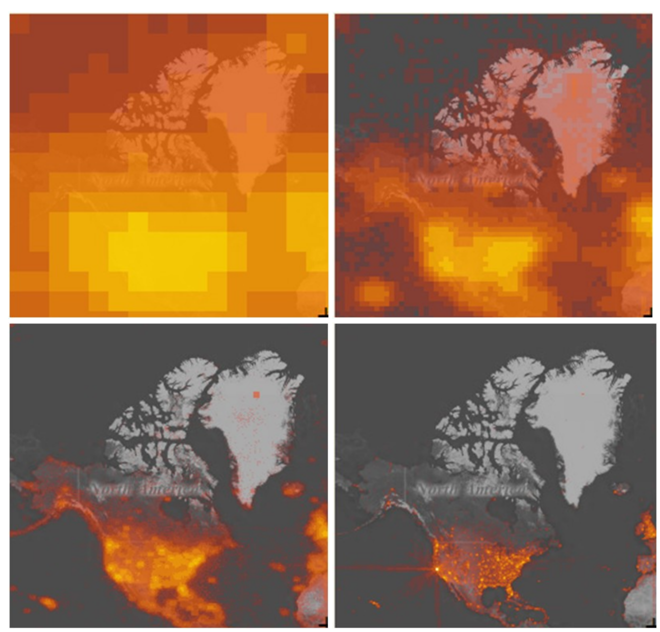

Figure 8. Increasing offsets for level of detail. One tile of Hotmap data at offset 4, 6, 8, and 11 from the original (level 1) tile.

driving directions (level 14 and 15) or cross-country driving (levels 6-7). The color map is adjusted for each zoom level, but stays consistent within the level.

The interface allows users to slice the data along several dimensions. Hotmap also allows users to specify a range of dates to examine (**B**). (It does not include an animation slider due to performance constraints in animating over the web.) Users can choose to focus on only one type of imagery, or can switch between them. The "style" selector (**C** in Figure 7) allows a user to specify results of road, aerial, hybrid imagery, or all together.

6 Design Decisions

Hotmap was designed after interaction with various customer groups, and was iterated during its deployment. It was incrementally improved and adapted to their needs. This section discusses some of the major decisions that went into building the interface.

6.1.1 Logarithmic Color Scaling

Tile requests are very unevenly distributed across the world: some parts of the world are looked at frequently, while others are hardly touched at all. The distribution is one familiar to researchers in social phenomena: a so-called "power-law" curve. When data is power-law distributed, a linear color scale will crowd at the bottom: very few sites will have a fraction of the hits of the top few. A logarithmic color scale, by contrast, distributes the lower values; finer distinctions in low levels are brought out. Figure 9 shows the distribution of frequency of tile hits at levels 13, plotted in log-log space. Note that over ten million road tiles received only one hit in our sample; a handful received over a million hits. The mapping from number of hits to color scheme varies by level of detail, but is kept constant within styles.

The logarithmic color scale helps increase the variance in color, but does not change the fact that most of the world's area would still be colored with the cooler end of the scale. In order to increase the visual salience of prominent places, we increase translucency as popularity diminishes. This is useful in order to figure out where most people are looking: very popular places turn vividly white or yellow, while less-popular places turn to little more than a faded highlight over the surface. This translucency is visible in Figure 8, for example, where the map can be partially seen below the image.

6.1.2 Catching Low Values

Sometimes, however, it is desirable to see whether particular obscure spots have been seen at all, as in section 4.2.5, above. Certainly, it provides useful feedback if no one at all has looked at a particular place. By disabling the translucency mentioned above, Hotmap highlights all points that anyone has looked. This leaves a version of the image colored in red where even one download has occurred, as seen in Figures 1, 3, and 5. As the log files are a sample of the servers, this is precise to all hits, but can usefully suggest the bounds of behavior. For broadly distributed behavior—like the scraping attack above—we can look for broad, but infrequent behavior. It can also be used to learn when users start to look at an area's imagery after it becomes available.

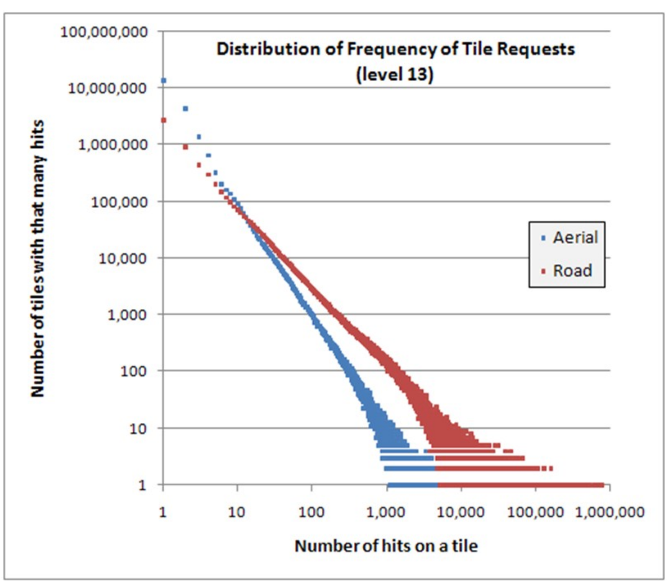

Figure 9. Frequency charts of tile use. On the *x* axis is a number of hits; on the *y* axis is the number of tiles with that number of hits, both plotted on a log/log scale for level 13 aerial (*blue)* and road (*red*) data.

6.1.3 Image Preview

Users frequently found it difficult to know exactly what tiles are visualized in the mash-up: it can be hard to remember the scale of an arbitrary level's data. Hotmap now has a second map, this one consisting a small preview pane, not unlike the inset on Figure 3, to show both the relative size of a single tile and the quality of imagery available at that point. The small map is linked to the main map control, and so always reflects the imagery at the center of the main map screen.

6.1.4 Using Mash-ups

The mash-up structure provided gained a number of advantages. In addition to the web-based deployment mentioned above, this visualization was able to leverage users' skills in navigating interactive maps. Also, because tiles are computed individually, the system was able to scale the visualization to arbitrarily large displays with no additional work. It is particularly interesting to run Hotmap on a very-large-screen display to examine large areas (Figure 10).

Mash-ups tend to overlay data over original maps. However, the map is optimized as a visual in itself and not a background; users reported that the high contrast background map was distracting. Hotmap enhances the look of the data and de-emphasizes the imagery by running a proxy that recolors Virtual Earth tiles as a low-contrast grayscale image. The range of grays that it offers do not overlap with the data overlay, which is a spectrum of reds through a vivid white. As a result, it is always clear which points contain data.

7 Other Applications

While this paper has focused on understanding user behavior and imagery acquisition, it is clear from the results that many different sorts of phenomena are visible in the system. This section suggests several different tasks, both server- and user-side, that may be accomplished with Hotmap, in addition to the ones alluded to above.

7.1 Server-Side Tasks

We have already discussed Hotmap's implications as system security dashboard, and the value of the tool to catch usability flaws. The power-law tile distribution mentioned above suggests that designers might optimize their tile delivery system for such an imbalanced load. For example, it would be wise to keep the most-requested tiles available in caches, preferably close to the user; the fact that the

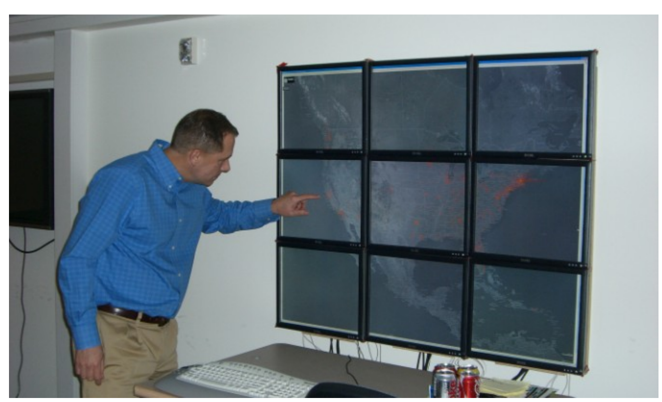

Figure 10. Displaying Hotmap on a Large Screen. A user studies Hotmap on a large 9-panel display.

distribution is so skewed might suggest that the cache need not be particularly large in order to be effective. (A back-of-the-envelope calculation suggests that eight out of ten tile requests could be handled within ten gigabytes of space, and nine out of ten could be handled within a hundred.)

There are advertising-relevant aspects of this information, too. It is now common for search engine advertisers to be able to know how many views they can expect from their advertisement. As geographical systems get locality-specific advertising, advertisers will be able to know how many people are likely to look in a given area, and thus whether their search term is relevant.

7.2 Automated Tour Guide

As noted above, even within well-populated areas, tourist attractions stand out as particularly popular. The wisdom of that crowd can be used to generate an "automated tour guide." This is social navigation in a literal sense: people's past movements become recommendations. The "footprints in the snow" [22] are written on the server log. We might use it either as an introduction to a new city—an enhanced map showing what places are the most looked-at—or as a cue for users wearing a mobile device.

7.3 Demography and Geography

Demographers and geographers have a particular interest in both how users approach the internet, and where their interest is focused. Zook [23], for example, examines the distribution of domain names per city around the world, looking to understand how internet content production compares to population; others have attempted to link internet use to geography through phone-based surveys. To the extent that usage of online mapping can be understood as a proxy for participation online—and remembering the principle that "I can see my house from here" turns the tool into an informal census—the maps produced by Hotmap can give a sense for Internet participation. It is worth reviewing the less prominent spots in Figure 2 to see where users are *not* looking.

8 Conclusion

Hotmap is an exploratory data visualization system, which takes advantage of the structure of the underlying data set to visualize it in its own space. In this case, we have examined geospatial imagery, which has the virtue of remaining fixed. When data sets are arranged spatially for users, linking their usage data back can provide important insights into how users approach the data. This is common in eye-tracking studies, such as [4], in which a heatmap is drawn over a web page to show where users looked first.

As an exploratory system, it helps to attack ill-formed problems. There is no set formula for imagery acquisition; thus, allowing users to reflect on the data from a variety of perspectives allows them to understand and explore the dataset in more depth. Hotmap provides a social view of the world, mapping a rough notion of popularity. This social view lets us see where users find the world to be interesting: what parts of it are worth a closer look? This will allow analysts, researchers, and designers to better understand how users interact with their world.

Acknowledgements

Hotmap was a highly collaborative effort. The author wishes to thank Paul Johns for database assistance; Sumit Basu for statistical assistance; Greg Strodel, Chad Raynor, and Steven Cosman for collecting data. Jon Howell, Jeremy Elson, and John Douceur implemented tile-layer mash-ups; and Shashikant Penumarthy and Jeremy Fisher provided valuable analysis. Shashikant also implemented the color map. I thank the members of the Microsoft Virtual Earth and VE Mobile team for feedback and promotion, and particularly Wolfgang Walcher, Brian Murphy, Joel Reiff, as well as the interviewees. Last, thanks to the members of the Community Technologies Group and VIBE at Microsoft Research, and particularly George Robertson, Jeff Heer, AJ Brush, Bongshin Lee, and the anonymous reviewers for valuable comments on the paper.

References

[1] T. Barclay, J. Gray, and D. Slutz. Microsoft TerraServer: A Spatial Data Warehouse. In *Proc. ACM SIGMOD* 2000. May 2000.

[2] E. Chi, P. Pirolli, J. Pikow. The Scent of a Site: A System for Analyzing and Predicting Information Scent, Usage, and Usability of a Web Site. In *Proc. ACM SIGCHI 2000*. April 2000.

[3] E. Chi, J. Pitkow, J. Mackinlay, P. Pirolli, R. Gossweiler, S. Card. Visualizing the Evolution of Web Ecologies. In *Proc ACM CHI 98*. May 1998.

[4] E. Cutrell & Z. Guan. What are you looking for? An eye-tracking study of information usage in Web Search. In *Proc ACM CHI 07*. May 2007.

[5] A. Dieberger, P. Lonnqvist. Visualizing Interaction History on a Collaborative Web Server. In *Proc ACM Hypertext 2000*.

[6] J. Elson, J. Howell, and J. Douceur. MapCruncher: Integrating the World's Geographic Information. *Operating Systems Review* 41:2. April 2007.

[7] I. Floyd, M. Jones, D. Rathi, M. Twidale. Web Mash-ups and Patchwork Prototyping: User-driven technological innovation with Web 2.0 and Open Source Software. In *Proc. IEEE HICSS 2007*. Jan. 2007.

[8] F. Girardin. "Heat Map of Barcelona Tagged Images." http://www.girardin.org/fabien/blog/2006/12/01/heat-map-of-barcelona-geotagged-images/ Retrieved March 24, 2007.

[9] J. Hong, J. Heer, S. Waterson, and J. Landay. WebQuilt: A Proxy-based Approach to Remote Web Usability Testing. In *ACM TOIS*. 19:3, July 2001.

[10] R. Kohavi. Mining E-Commerce Data: The Good, the Bad, and the Ugly. In *Proc ACM KDD 01*. Aug 2001.

[11] K. Lynch. *The Image of the City*. Boston: MIT Press. 1960.

[12] D. Martin (1996) An assessment of surface and zonal models of population *International Journal of Geographical Information Systems* 10, 973-989

[13] A. Mehler, Y. Bao, X. Li, Y. Wang, and S. Skiena. (2006) "Spatial Analysis of News Sources." In *Proc Information Visualization 2006*.

[14] R. McGovern. MSDN: Building Your Own Tile Server. http://msdn2.microsoft.com/en-us/library/bb259690.aspx Retrieved March 24, 2007.

[15] T. Munzner. Visualizing the Structure of the World Wide Web in 3D Hyperbolic Space. In *Proc. ACM VRML '95*. January, 1995.

[16] S. Murthy, D. Maier, L. Delcambre. Mash-o-matic. In *Proc ACM DocEng '06*. Oct. '06.

[17] J. Oates. "Microsoft's Earth deletes Apple HQ". *The Register*. July 25, 2005. http://www.theregister.co.uk/2005/07/25/msn_earth_deletes_aple/ Retrieved March 24, 2007.

[18] A. Sen, P. Dacin, and C. Pattichis. Current Trends in Web Data Analysis. In *Communications of ACM* 49:11, Nov 2006.

[19] T. Slocum, R. McMaster, F. Kessler, and H. Howard. *Thematic Cartography and Visualization*. New Jersey: Pearson Prentice 2005

[20] US Census Bureau. "Census 2000 Population Distribution in the United States", Sep. 2001. http://www.census.gov/geo/www/mapGallery/2kpopden.html. Retrieved March 24, 2007

[21] M. Wattenberg. Baby Names, Visualization, and Social Data Analysis. In *Proc IEEE Infovis 2005*. October, 2005.

[22] A. Wexelblat and P. Maes. Footprints: History Rich Tools for Information Foraging. In *Proc. ACM CHI 1999*. May 1999.

[23] W. H. Whyte. *City: Rediscovering the Center*. New York: Anchor. 1990.

[24] M. Zook. Old Hierarchies or New Networks of Centrality? The Global Geography of Internet Content Market. *American Behavioral Scientist* 44:10. 2001.