

# The Computational Geowiki: What, Why, and How

Reid Priedhorsky and Loren Terveen  
GroupLens Research  
Department of Computer Science and Engineering  
University of Minnesota  
Minneapolis, Minnesota, USA  
{reid,terveen}@cs.umn.edu

## ABSTRACT

Google Maps and its spin-offs are highly successful, but they have a major limitation: users see only *pictures* of geographic data. These data are inaccessible except by limited vendor-defined APIs, and associated user data are weakly linked to them. But some applications require access, specifically *geowikis* and *computational geowikis*. We present the design and implementation of a computational geowiki. We also show empirically that both geowiki and computational geowiki features are necessary for a representative domain, bicycling, because (a) cyclists have useful knowledge unavailable except from cyclists and (b) cyclist-oriented automatic route-finding is enhanced by user input. Finally, we derive design implications: for example, user contributions presented within a route description are useful, and wikis should support contribution of opinion as well as fact.

## Author Keywords

Wiki, geowiki, computational geowiki, web-map.

## ACM Classification Keywords

H.5.3 [Group and Organization Interfaces]: Collaborative computing, computer-supported cooperative work, web-based interaction.

## INTRODUCTION

Google Maps transformed mapping, bringing geographic information systems to the masses in a convenient form. “Mashups”, visualizations of existing data sets on top of Google Maps, soon emerged – ranging from estimating cab fare<sup>1</sup> to finding volcanoes<sup>2</sup> to sending “geogreetings”<sup>3</sup>. Next came open content – sites like WikiMapia and PlaceOpedia added wiki ideas, letting anyone add or edit places, text, and other information on top of a map.

<sup>1</sup><http://yellowcabnyc.com/farefinder>

<sup>2</sup><http://geocodezip.com/volcanoBrowser.asp>

<sup>3</sup><http://geogreeting.com/>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CSCW’08, November 8–12, 2008, San Diego, California, USA.

Copyright 2008 ACM 978-1-60558-007-4/08/11...\$5.00.

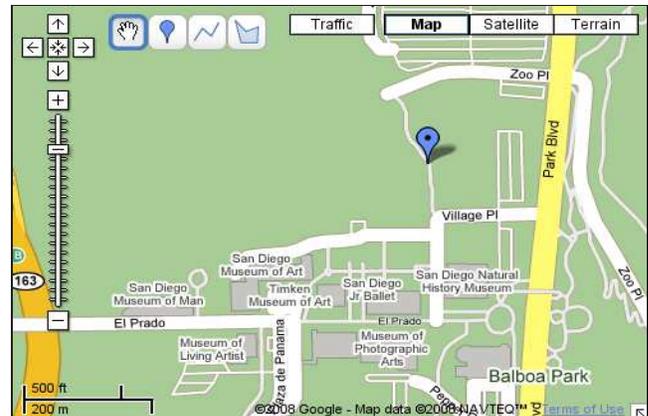


Figure 1. Google My Maps view of the San Diego Zoo and vicinity.

However, the powerful Google Maps infrastructure has a major limitation: users cannot interact with the geographic *data*. Instead, they are shown *pictures* of the data, and any content they add is a separate – and weakly linked – layer.

Consider the Google My Maps view of San Diego shown in Figure 1. The “pushpin” near the center indicates a user-added place. Users can collaboratively add and edit arbitrary descriptions of this place. However, contrast this with the various museums filling the lower half of the map – users cannot share their experiences at (e.g.) the Timken Museum of Art, because its representation is just a set of pixels.

To be precise, Google knows about these places, but it does not expose them as interactive objects unless and until a user searches for them. Then they show up as pushpins, and users can add them to a My Map, read or write descriptions, or do other place-related interactions. However, the possibilities are only those that Google has chosen to provide. In particular, the possibilities for the transportation network are very limited. Users can’t even share information like “avoid this section of Florida Drive during rush hour”, let alone implement a new route-finding algorithm.

This matters because some applications *require* access to this geographic data. A full-fledged wiki and many promising computational applications are simply impossible on top of Google Maps. More deeply, varying levels of geographic data access enable different types of applications:

1. **Pictures.** Geographic data is available only through pre-rendered image tiles or specialized API functions such as geocoding or computing driving directions. That is, one can only “see” the map data; one can’t interact with it except in limited vendor-chosen ways. As noted above, maps built on APIs provided by Google and its peers provide only this level of access to base map data.
2. **Interaction.** Map data is available as individual objects and relationships but cannot be modified. This is required for applications needing to *compute* using geodata – e.g., to implement a custom route-finding engine – or simply to annotate (tag, rate, comment on) geographic objects without relying on weak co-location links. Such applications include snow removal (a citizen could notify the city that “*this* alley is icy”) and wildfire management (incident commanders could compute risk scores for fire-fighting teams based on fire threats to and other attributes of their designated evacuation routes).
3. **Modification.** Map data is available as objects and relationships and can be modified. This is required to implement a full-fledged *geowiki*, where users can add, edit, and monitor *all* geodata without arbitrary distinction between “user” and “base” data.

Of particular note are computational geographic applications that require user input – *computational geowikis*. For example, bicyclists need route finding in a transportation network whose properties and structure are fully known only by the cycling community, and which may change [27]. Many transportation domains have similar problems. Another important domain is public health: for example, effective modeling of epidemic progression depends on accurately capturing the geographic behavior of carriers – information collected in a distributed fashion by individual doctors.

Our work presents the following contributions. We invented the notion of a computational geowiki and created such a system, producing significant design innovations and solving major implementation challenges. To evaluate our system, we studied it in the context of the bicycling community (chosen for reasons detailed later). We show empirically that cyclists know much cycling-relevant information that is useful and not available except from cyclists; thus, a geowiki is a powerful knowledge sharing tool for the cycling community. We also found that the crucial computational service of route-finding may be more accurate when based on user input; thus, a computational geowiki offers enhanced value to cyclists. Our results in this representative domain suggest the general utility of geowikis and computational geowikis.

In the remainder of the paper, we first survey related work, focusing on existing systems that instantiate aspects of the geowiki approach, then briefly describe the bicycling domain. We next describe the design and implementation of our geowiki, illustrating the innovative design features and explaining key implementation challenges solved. We then present our experimental design and methods, followed by our results and implications. We conclude with a summary and possibilities for future work.

## RELATED WORK

Over the past decade, a new form of collaborative interaction has emerged on the web. Users no longer just consume information or discuss topics; they work together to produce *artifacts of lasting value* [11]. Collaborative filtering systems leverage users’ ratings of items (movies, books, consumer products, etc.) to enable personalized recommendations. News sites like Reddit, Digg, and Slashdot rely on user opinions to filter and order stories. Tagging systems like del.icio.us, Flickr, and CiteULike let users associate keywords with items, facilitating searching and exploration. Q&A sites like Yahoo! Answers and Experts-Exchange form knowledge economies, where users spend points to ask questions and earn them for answering. Finally, wikis take the user-provided content to its logical end: anyone can add, edit, or delete anything.

Scholarly interest in this form of interaction is intense, encompassing both studies of current sites and techniques and efforts to develop novel and better ones [6, 12, 17, 22, 30].

### Wikis

The Wikipedia online encyclopedia is the most famous and successful wiki, containing over 2 million articles in English, ranking among the 10 most visited sites on the Web, and possibly having accuracy similar to *Encyclopedia Britannica*, at least for some types of articles [15]. Much research has been devoted to Wikipedia. For example, Viégas et al. [32] and Priedhorsky et al. [26] have shown that vandalism and other damage in Wikipedia is usually repaired quickly. Kittur et al. [20] and Priedhorsky et al. [26] found that small minorities of editors produce the vast majority of the content and value of Wikipedia. Adler and Alfaro developed a scheme to compute reputation of authors based on whether their edits persisted or were removed, and used this metric to compute the credibility of each section of an article [1]. Bryant et al. interviewed Wikipedia editors and described how the activities they engage in on the site change as they progress from novices to experienced editors [8]. Finally, theoretical work by Cosley et al. suggests that users are as effective as experts in reviewing other users’ work [10], and that the wiki model and traditional review-before-publication result in the same quality, but the wiki model achieves it faster [11].

### Geography

The field of geographic information systems (GIS) is concerned with visualizing, analyzing, and manipulating geographic and spatial data [23]; traditionally, GIS work is done by experts using specialized software. The GIS community has proposed various types of collaboration [2, 24, 29].

In particular, geographic “citizen science” is becoming an established approach. For example, an annual bird census is done mostly by laypeople [7], and the National Map Corps is a volunteer program to correct and update United States Geological Survey maps [4]. Professional geographers manage these projects and process and vet the data submitted by citizens; often, as in the case of the National Map Corps, these professionals are a bottleneck [4].

More recently, geographers have become interested in what they call “volunteered geographic information”. Goodchild has argued for the value of average people’s geoknowledge [16], and geographers held a scholarly meeting in 2007 to set a research agenda for the area<sup>4</sup>. However, no systems have been built, and geographers distrust the wiki model, perceiving tension between data quality and open content [16]. On the other hand, open content is ascendant within the CSCW community. This is due in no small part to the great success of Wikipedia – contrasting dramatically with the failure of its predecessor, Nupedia, which had an elaborate credentials-based review process.

### Steps toward Geowikis

Adapting a wiki to the geographic domain results in a *geowiki* [27]. Key adaptations include: WYSIWYG editing to make manipulation of geographic objects tractable for users; versioning that works on a landscape of tightly coupled geo-objects rather than a collection of individual documents; and monitoring tools appropriate for the geographic domain, such as *watch regions* (rather than watch lists) and visual diffing tools to show geographic changes effectively.

Many systems partially implement the geowiki model. OpenGuides is a wiki travel guide, WikiMapia lets users enter and edit information for places and rectangular regions, and Google Maps lets users edit the locations of searched-for places and add new places. Google My Maps goes beyond this, allowing collaborative editing of geographic points, paths, and polygons, all of which can be annotated with text, images, and videos. “Digital graffiti” systems let users associate virtual information with physical geography, and this information can then be accessed *in situ* with location-aware devices [9, 14]. Fundamentally, however, these systems work at the Pictures level: while they put wiki information into a geographic context, the base geodata are not available for annotation or editing.

Other projects come closer to true geowikis. The most famous is Open Street Map<sup>5</sup>, an effort to build a worldwide street map with a wiki. Another is the Cambridge Cycling Campaign’s “journey planner” for cyclists<sup>6</sup>. These projects offer WYSIWYG editing of the transportation network, but key wiki features like watch regions and a recent changes list are missing<sup>7</sup> – meaning that the transparency necessary for effective collaborative editing is missing. Also, they focus solely on mapping the transportation network, so annotations and other user contributions beyond editing the network are not possible. These sites offer some functionality at the Modification level but none at the Interaction level.

Efforts to enhance navigation with user contributions also exist. Some GPS devices let users enter map corrections directly into the device and subscribe to corrections made by other users [31]. Navteq, a major mapping firm, accepts

<sup>4</sup><http://www.ncgia.ucsb.edu/projects/vgi/>

<sup>5</sup><http://www.openstreetmap.org>

<sup>6</sup><http://www.camcycle.org.uk/map/route/>

<sup>7</sup>While a recent changes list is specified in the OSM API, it is not implemented in the OSM editing tools.

user map corrections on its website, but these are checked by professionals before being offered to other users [25]. Bederson et al. have proposed a system for augmenting automobile route-finding with “subjective human experience” [5]. Finally, Counts and Smith have proposed a system that uses location-aware sensors to construct route information for individuals that then can be shared with others [13].

Previous work has demonstrated the utility of open content, in particular wiki-style mass collaborative editing. However, neither scholars nor practitioners have previously extended this notion to geography in the comprehensive way essential to many applications.

### THE CYCLING DOMAIN

Prior work suggests that cycling is a good domain for a geowiki. We studied the information needs of the bicycling community [27], finding that cyclists had a tradition of knowledge-sharing, lacked an up-to-date and comprehensive resource for obtaining cycling-oriented information and routes, and were enthusiastic about the wiki model. The current paper gives empirical evidence for some beliefs held by cyclists, which were reported in this past work.

Many existing web sites aim to support cyclists’ route information needs. Gmaps Pedometer<sup>8</sup>, Bikely<sup>9</sup>, and others let users manually define and share routes. Wayfaring<sup>10</sup> enables collaborative editing of routes and places of interest. A few offer automatic route finding, e.g., YTV Journey Planner for Cycling<sup>11</sup> and Fietsrouteplanner<sup>12</sup>. Closest to our work is the Cambridge Journey Planner (noted above), which offers user editing of the transportation network but lacks critical wiki features. Again, these sites recognize the potential of open content to meet the information needs of cyclists, but are limited by incomplete scope or clumsy user interfaces.

Our system provides cyclist-specific automatic route-finding. It uses the A\* search algorithm [18] over the transportation graph, weighting edges (blocks) with length times bikeability. We computed bikeability using two methods. If enough attributes were available for a block (at least speed limit, daily traffic volume, and number of lanes), we used a metric developed by the Chicagoland Bicycle Federation [3] (here called CBF), with minor modifications<sup>13</sup>. We chose this metric because it was the only existing metric we could find whose input parameters were covered in the data for our area (which has typical coverage). Even so, about 20% of the blocks in our database did not have enough data to compute CBF. For these, we used an ad-hoc metric we call

<sup>8</sup><http://gmap-pedometer.com>

<sup>9</sup><http://www.bikely.com>

<sup>10</sup><http://wayfaring.com>

<sup>11</sup><http://kevytliikenne.ytv.fi/?lang=en>

<sup>12</sup><http://fietsersbond.nl/fietsrouteplanner/>

<sup>13</sup>Specifically: expressways were always rated very poor; wide shoulders earned a 1.5-star rating increase, not 2, and very wide shoulders did not earn an additional increase; the presence of bike lanes increased the rating by 0.5 stars, and bike lane width was included in shoulder width; all lanes were assumed to be 12 feet wide. These modifications were motivated by data availability and the first author’s expertise as a bicyclist.

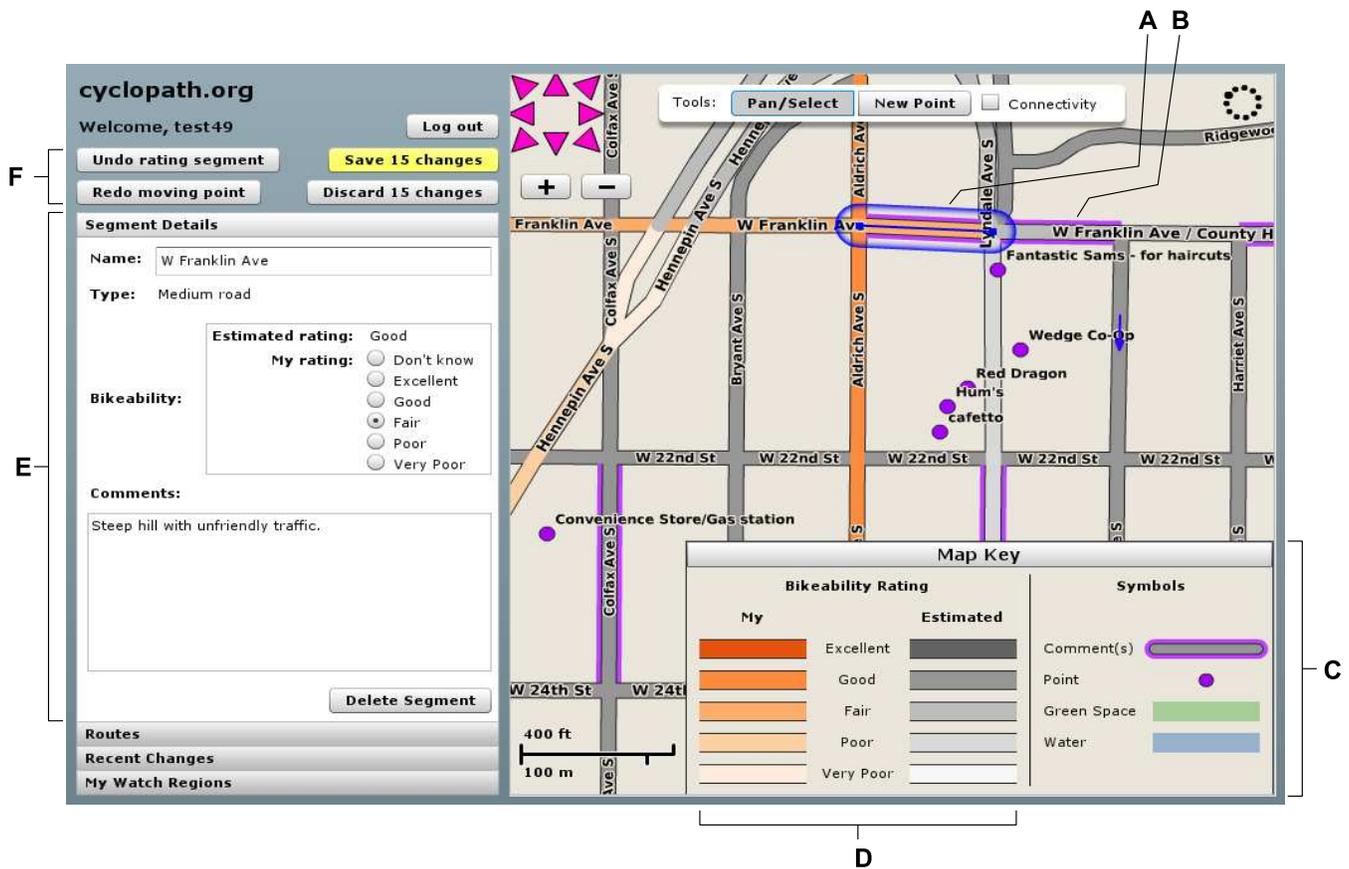


Figure 2. Screenshot of our user interface. On the right is a modern “slippy” map which works similarly to Google Maps, while on the left are details about the selected object. User-contributed content, including points and annotations on blocks, is visible. (This figure is best viewed in color.)

“Naive”: roads were rated based on their type (e.g., municipal street, state highway, U.S. highway), and bike paths were rated based on the length of the block (longer distances between intersections were rated better).

## GEOWIKI DESIGN AND ARCHITECTURE

Our fundamental challenge was to produce a system that rivaled Google Maps in quality of user experience and performance, while also offering full implementations of Interaction- and Modification-level geowiki functionality. In this section, we outline the main design features<sup>14</sup> (with reference to the screenshot shown in Figure 2), then identify implementation challenges and how we solved them.

### Design

**Web-map interaction.** Basic display and navigation work similarly to Google Maps: geographic features such as streets, water bodies, and green space are distinguished visually using color, and users navigate by dragging the map or using the pan/zoom controls. A hideable map key (C) reminds the user how colors and other marks map to data.

**Comments and Ratings.** Users can click on any geographic object (e.g., the road segment A) to bring up an editor (E) to

view and edit the object’s attributes, comments, and rating. The meaning of a rating is domain-specific: here, the rating of a *block* (i.e., a segment of road from one intersection to the next) indicates its bikeability on a scale of 1 (“very poor”) to 5 (“excellent”). For blocks not yet rated by the user, we estimate ratings using techniques explained below. Block annotations are visualized on the map; each block is colored to indicate its rating and the source of that rating (D), and a purple halo (e.g., B) indicates comments on it.

**Wiki Editing.** Comments on geographic objects can be edited by any user. Further, geographic objects themselves can be edited; any user can add, delete, or modify the geometry of points (places) or blocks. Ratings, however, are private to each user; this follows standard practice. As in text wikis, in-progress work is managed on the client before being saved in batch to the server upon user command, and users have access to unlimited undo and redo.

**Wiki Monitoring.** Again as in text wikis, users can review the editing activity of other users. A geowiki provides a history browser (recent changes list) and *watch regions*, which adapt the wiki watch list to the geographic domain. Users can define regions of interest and then be notified when geographic objects in those regions are modified; these regions can be defined as polygons or relative to a route (e.g., “notify me of any changes within 200 meters of my commuting

<sup>14</sup>As of this writing, a few features are unimplemented, specifically, rating places, watch regions based on routes, and reverting.

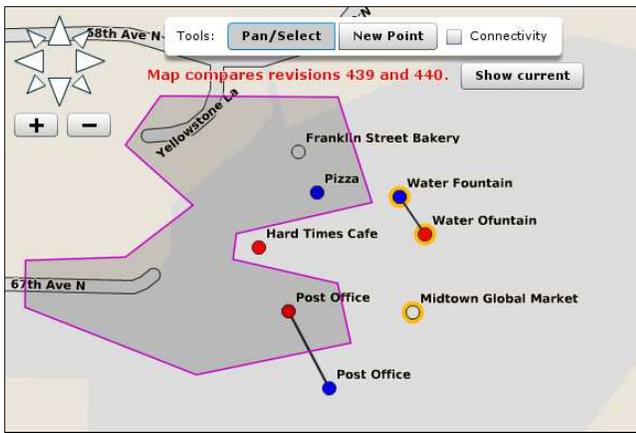


Figure 3. Geographic diffing. New geometry is blue, while old geometry is red; if a point was moved, the old and new versions are connected with a line. An orange halo indicates non-geographic changes (“Midtown Global Market” and “Water Fountain”). The user would be notified that “Pizza” was added and “Post Office” was moved, since these changes intersect his or her (mocked-up) watch region.

route”). Also, *geographic diffing* (Figure 3) helps users visualize changes in geography.

**Computational Support.** Finally, the geowiki supports domain-specific algorithms. One useful in many domains is route finding: users enter start and end addresses and perhaps domain-specific preferences such as block properties that are important to them. The system computes a route and shows it both visually on the map and as a printable list of turns. As noted above, we compute routes using the A\* algorithm, with a domain-specific edge weighting function. Currently, we use generic bikeability metrics. However, our results (detailed below) argue for personalizing weights using a technique like collaborative filtering [21].

### Implementation

The fundamental implementation challenge in realizing our design was to create a web-based system with a good user interface and acceptable performance. This section details these challenges and how we solved them.

**Performance.** Achieving good performance for a geowiki is difficult, more so than for Google Maps and other systems that operate at the Pictures geodata access level. Not only must a web browser manage thousands of interactive, clickable objects, geodata must be delivered from the server to the client in a form manipulable by users and software. In other words, it seems impossible to serve pre-rendered image tiles.

However, we make a critical observation: in practice, geographic data objects cannot be manipulated by users if they are too “small” visually. In other words, if a map is zoomed out, users can’t visually distinguish or accurately point to specific geographic features. We exploit this observation by implementing a two-part scheme for serving geodata. When zoomed out beyond a map scale of about 1:24,000 (a region roughly 4km square in a 1024x768-pixel window), we serve pre-rendered tiles; when the user zooms in, reducing

the number of geo-objects in the viewport to a tractable level, we switch to serving individual objects containing vector geometry and full attributes.

The performance savings gained by this scheme are significant. For example, at the most-zoomed-in level where we serve raster tiles, a typical view of an urban area might contain 3-4,000 geographic objects comprising about 250 kB of either compressed vectors or raster tiles. However, the interface would be unacceptably slow if it had to render this many discrete objects, and when zooming out  $n$  levels, space consumption increases by  $O(n^2)$  in vector mode but is roughly constant in raster mode.

**Client technology.** We implemented the web client in Adobe Flex; it runs in a browser under the widely deployed Flash Player 9 virtual machine. We evaluated other technologies, notably AJAX and SVG, but found them too slow or not widely supported; also, the user experience quality was significantly better with Flex. The client communicates with the server using HTTP, transferring geo-objects and support data using a lightweight custom XML serialization protocol and tiles using standard HTTP file requests.

**Versioning.** Wikis require versioning: the system stores previous states, making it possible and easy for users to understand the progress of work and to *revert* undesired edits – both critical to making the open wiki editing model work. In a geowiki, inter-object relationships (e.g., which blocks connect to which other blocks) are as important as the objects themselves, so versioning operates globally rather than at the per-article level typical of a text wiki. The history browser can “filter” revisions, displaying only those affecting objects in the current viewport or a specified watch region.

**Base geodata.** We seeded the wiki with GIS data from our state’s Department of Transportation. These data are free to obtain, modify, and republish; this is typical for the United States but unusual in many parts of the the world (the motivation for Open Street Map). Considerable manual cleanup of the data is required, a task well-suited for the wiki model.

**Status.** We are in the final stages of alpha testing, with about 100 users, and preparing to deploy the system for public use.

### EXPERIMENTAL DESIGN AND METHODS

As mentioned above, we seek to establish when *geowikis* are useful and when *computational geowikis* are useful, general questions which cannot be studied in general. We address them with an experimental evaluation of our computational geowiki applied to the cycling domain; results using this representative community also provide evidence at the general level. This section explains our recruiting methods and experimental design.

**Subjects and recruiting.** We recruited active cyclists using online methods, posting invitations on several mailing lists and forums frequented by cyclists as well as Craigslist. This skewed our subject pool towards cyclists who are comfortable on computers, which seems reasonable because we are

evaluating computational support tools. Subjects were 18 or older and had spent at least 3 hours or 25 miles bicycling in the Uptown neighborhood of Minneapolis, Minnesota during the year preceding the study. They were compensated with a gift certificate to a local bike shop.

After conducting 6 pilot interviews, we ran the experiment with 30 subjects. One of the interviews failed, so we report results for 29 subjects. 17 of the subjects were men and 12 women. 17 reported riding “nearly every day”, and 19 rode at least 50 miles per week.

**Tasks.** The basic form of an interview was a series of four tasks, each followed by a mini semi-structured interview. The four tasks were:

- T1. **Route-finding.** The subject used our system to compute a familiar route of their choosing. *We asked:* whether the subject liked or disliked the route, how they currently learned new routes, and how useful these learning methods were (on a scale of 1 to 5).
- T2. **Entering places.** The subject entered at least 4 new points (places) that he or she wanted to share with other cyclists. *We asked:* why the new places were chosen, whether it was difficult to think of places to enter, how subjects currently found information about cycling-relevant places, and how well these information-finding methods worked.
- T3. **Editing comments.** The subject entered or modified at least 4 comments about places or blocks. *We asked* questions analogous to Task 2. Additionally, when any of the last 17 subjects read an existing comment, we asked them if it was useful.
- T4. **Rating bikeability.** The subject rated the bikeability of 12 or more blocks on a 5-star scale. *We asked* questions analogous to Task 2.

Points and comments edited by one subject were visible to subsequent subjects, while routes and ratings were not. Keeping routes and ratings private is consistent with standard practice, since this information is considered personal. We concluded interviews by asking questions about privacy, the usefulness of several different information resources, and satisfaction with current information-gathering methods.

## RESULTS

We next present our results, which show the utility of a geowiki and computational geowiki for cyclists. First, cyclists knew useful information that was not available otherwise. Second, route-finding may be more accurate when based on user input than when based on standard bikeability metrics.

### R1. Cyclist knowledge useful and unavailable elsewhere

Subjects had no problem entering the data they were asked to, often entered additional data, and reported that it was easy to think of this data. Further, our content analysis revealed a rich diversity of entered information – “subcultural community” resources, personal experiences and advice, and detailed cycling-relevant information. Finally, this

information is clearly useful to cyclists and difficult to obtain except from cyclists.

### R1a. Cyclists have knowledge

**Summary statistics.** *Places:* Subjects entered a total of 129 new places. 28 of 29 subjects entered at least the 4 requested places, and nine entered at least 5. *Comments:* Subjects edited a total of 224 comments (71 on places, 153 on blocks), with 32 edited by at least two subjects (19 on places, 13 on blocks). All subjects edited at least the 5 requested comments, and seven edited 12 or more. *Ratings:* Subjects entered 828 bikeability ratings for blocks. 26 subjects entered at least the 13 requested ratings, nine entered at least 23, and the top two entered 88 and 121 ratings.

**Knowledge self-reports.** A majority of subjects reported that it was easy to think of comments (25 of 29) and ratings (17); 10 subjects reported ease in thinking of new places. However, majorities in all three tasks thought that it would be easy if they were not under the pressure of an interview (16, 26, and 19 subjects on places, comments, and ratings, respectively) and that they had more information they could share (24, 22, and 20).

### Content analysis.

The core of our argument that cyclists have knowledge is a content analysis of the places (Task 2), comments on places (Task 3), and comments on blocks (Task 3) entered by subjects. We coded these three groups of data separately, considering both the data themselves and relevant interview notes. Two coders independently defined categories for each group, then met to produce consensus categories. The same two coders then independently applied these categories, resolving disagreements by discussion. In general, categories were non-exclusive – i.e., one item can be in multiple categories at each level – except for the *Other* categories. Therefore, membership counts cannot be summed.

**Places.** Table 1 summarizes our content analysis of places. We found three major categories. *Lifestyle/community* places relate to everyday urban life, e.g. post offices, zoos, bookstores, parks, and art galleries. *Food/drink* places are where these two items can be obtained. *Cycling-specific* places are directly related to the practice of bicycling, including landmarks, shortcuts, meeting places, and big hills. A surprising observation is that few places are cycling-specific: this category accounts for only 23% of the total. We first discuss these places, then the other two categories.

The largest sub-category of cycling-specific places is *landmarks*. In principle, nearly any place could serve as a landmark; however, we categorized a place as such only if a subject described it that way explicitly (e.g., to “key off of”) or if this use was clear from their interview comments. *Meeting spots* are analogous. *Road/trail* information would ideally have been attached to a block; that is, if subjects had understood the system fully, these would have been block comments instead of place comments. Some of these cycling-specific places strikingly illustrate our hypothesis that the knowledge cyclists have is difficult to obtain; e.g., that cy-

Category	#	Example(s)
<i>Lifestyle/community</i>	53	
Community resource	22	post office ; Como Zoo ; Walker library ; YWCA – uptown
Retail – non-food	17	Arise Bookstore ; Target ; Chicago Lake Liquor
Park	14	Hidden Beach ; Como Zoo ; Community Garden
Arts	14	Orchestra Hall ; Soap Factory [an art gallery]
<i>Food/drink</i>	47	
Restaurant	23	Black Forest restaurant ; Longfellow Grille
Coffee shop	17	Espresso Royale
Groceries	8	Kowalski’s ; Byerly’s ; Penzie’s Spices
Water fountain	2	water fountain
<i>Cycling-specific</i>	30	
Landmark (explicit)	13	Obnoxious Billboard ; Lighthouse ; Old Grain Belt Brewery
Road/trail	8	Brackett Park, Bridge under greenway to 37th Ave. ; New Greenway Bridge!
Cycling-specific (other)	7	franklin ave hill ; difficult intersection to ride
Meeting spot (explicit)	3	Critical Mass gathering area ; bike trick hang out
<i>Other</i>	6	

Table 1. Categorization of the 129 places entered by subjects.

Category	#	Example(s)
<i>Objective place information</i>	60	
Description	57	collectice [sic] bookstore with activist literature ; No eating spot inside, though.
Events	6	Starting point for Messenger Challenge and Stupor Bowl
<i>Subjective place information</i>	47	
Review	44	Wonderful Asian food; Best priced bike repairs
Advice	6	Lock your bike here, I know the people all seem cool but do it anyhow
Personal narrative	3	we enjoy visiting this place ; Every time I ride this street, I see something new.
<i>Cycling-specific</i>	23	
Bike parking	9	Plenty of bike parking in front. ; No bicycle-specific parking
Bikeability notes	8	Silly one way trails ; useful exit/entrance off/onto the Greenway
What cyclists do at place	7	Lock your bike here; Starting point for Messenger Challenge and Stupor Bowl
<i>Other</i>	3	

Table 2. Categorization of the 71 place comments edited by subjects.

clists gather at a particular place to do bike tricks, or that a particular lighthouse is a critical landmark on certain routes. Cycling-specific places can play an important role in route-finding; e.g., relevant ones – perhaps chosen in a personalized way – could be added to a route description to aid orientation and make it easier to identify turns.

Turning our attention to non-cycling-specific places (76% of the total), we assume that subjects followed task instructions, believing that fellow cyclists would find these places interesting even though they weren’t about cycling per se. Why would they think this? We conjecture that cyclists know their peers well enough to know what else – beyond cycling – they tend to like. The places they entered form a cultural snapshot of the local bicycling community; i.e., they mark information of interest *to cyclists*, but not necessarily of interest *while cycling*. This result, consistent with [28], suggests that information resources for cyclists, and perhaps for many other groups, should support off-topic conversation as a useful means to build community, rather than suppressing it as undesirable noise.

**Place comments.** The tendency to enter non-cycling-specific information continued for place comments, although not as

strongly; Table 2 summarizes these comments. The most popular categories were *objective place information* – essentially factual descriptions of places – and *subjective place information*, typically comprising brief free-form reviews. 92% of place comments fell into one of these two categories, while 32% contained cycling-specific information. (Note that there was considerably more overlap in categories for comments than for places.)

However, here the critical distinction is between objective and subjective information. This shows that geowikis must record *opinion* as well as fact. Some wikis already provide for non-factual discussion; for example, each Wikipedia article comes with an associated “talk” page where users can discuss the article’s content. However, these talk pages are for discussion of what the facts are and whether Wikipedia conventions are being followed – *not* the expression of subjective opinions about the subject matter. Our data suggest that objective fact and subjective opinion both deserve a first-class role. Therefore, we believe an organization similar to that of product reviews on e-commerce sites is more appropriate: both fact (features, price, etc.) and user opinion are present, and clearly distinguished. This result also shows the desirability of letting users rate places.

Category	#	Example(s)
<i>Description</i>	108	
Lane/facility type	61	pretty wide ; Bike lane on right side
Description (other)	40	Several road crossings with 4 way stop signs
Surface quality	22	smooth pavement ; a lot of potholes
Snow removal	10	Plow conditions are not that great here
Current conditions	6	not currently plowed well (12/10/07)
<i>Worries/annoyances</i>	100	
Motor traffic (quantity)	48	heavy traffic ; Quiet
Worries/annoyances (other)	39	many stop signs ; sketchy at night (crime)
Motor traffic (behavior)	20	people in a rush ; can be dangerous because of turning cars
Hazards	19	possible broken glass hazards ; covered in snow, very slick
Construction	9	Construction is ongoing ; Lots of construction in this area
<i>Subjective information</i>	41	
Advice on route choice	35	great place to enter the greenway coming from the whittier neighborhood.
Personal narrative	5	Just recently discovered
Scenery	3	Gets into some sections with woods and hills that surround the trail
<i>Other</i>	6	

Table 3. Categorization of the 153 block comments.

**Block Comments.** Comments on blocks are summarized in Table 3. The major categories are *descriptions*, including basic properties of a block like its width and surface type; *worries/annoyances*, e.g. quantity and quality of motor traffic or construction; and *subjective information*. In contrast to places and place comments, virtually all block comments relate directly to cycling. We conjecture that this is because when subjects focused on where they actually cycled – the paths and roadways – they naturally thought of information useful *while cycling*. Block comments are useful in both evaluating and following routes; e.g., a comment that a block has many potholes could lead a cyclist to choose a different route or to ride that route with better preparedness.

Intuitively, the type of information revealed by our content analysis seems useful: cyclists “obviously” want to know the places they can go, what those places are like, and what to expect on their way there. We next present our data, which support this intuition.

#### R1b. This knowledge is useful

Subjects found information from other cyclists useful, both when asked about specific entered information and when asked about the general utility of such information.

**Other cyclists’ comments were useful.** When a subject read a comment edited by another subject, we asked if the comment was useful (for the first 6 comments read per subject). 50 of 64 comments read (70%) were judged useful.

**Other cyclists were considered useful.** We asked subjects about the utility of various sources of cycling information. When queried after each task about current information sources, subjects frequently mentioned other cyclists, giving an average utility rating of between 3.5 and 3.9 out of 5, depending on the task. Also, in the final set of questions, subjects rated the utility of other cyclists’ bikeability opinions as a mean of 4.14. This is slightly higher than (though not statistically different from) their rating of “objective bikeabil-

ity factors” at 3.97. We suspect that these figures actually understate the utility of other cyclists’ opinions. 2/3 of the subjects said that individual cyclists differ in their bikeability assessments. Therefore, we conjecture that a system that computes *personalized* routes using only the ratings of like-minded cyclists will lead to other cyclists’ opinions having a higher perceived value.

#### Social connections may mean better knowledge access.

Many types of information flow through social networks; our results hint that this is true for cycling-related information as well. The concluding survey asked subjects to respond to the following two statements on a 5-point Likert scale, from “strongly disagree” to “strongly agree”: (1) “It is hard to find other cyclists who have the specific information I need” and (2) “I am satisfied with my existing methods of gathering information for planning rides and selecting routes”. The correlation between responses to these two items was -0.40. In other words, having weak links with other cyclists may make it harder to find cycling knowledge. This argues for a geowiki, which collects and distributes knowledge and supports social ties.

#### R1c. This knowledge is available only from cyclists

After each task, we asked subjects how they currently obtain each type of information they entered into the system and how useful their methods were. Table 4 details the responses. Two key findings are apparent:

- “Trial and error” was most useful for every task (4.2 to 4.6 out of 5), and was mentioned second most often. In other words, subjects told us the best way to get useful knowledge was to go out find it themselves. While effective, this is time-consuming and particularly unhelpful when answers are required immediately.
- “Word of mouth” was mentioned most often and was rated as quite useful (3.5 to 3.9). That is, cyclists try to benefit from each other’s experience whenever they can. While

Source	Task 1	Task 2	Task 3	Task 4
Word of mouth	18 3.5	23 3.8	23 3.9	19 3.8
Trial & error	24 4.2	8 4.3	18 4.3	14 4.6
Internet forums	7 3.9	6 3.4	14 3.3	11 3.2
Paper maps	15 3.8	5 4.0	2 2.5	15 3.3
Online maps	17 3.6	3 4.3		
Internet search		11 4.2	4 3.5	
Specific websites		6 3.7	4 2.3	
Newspapers		4 2.9	2 3.8	
Phone book		6 2.3		
Advertisements		3 2.3	1 2.0	
Other	5 2.8	4 3.1	3 3.0	3 2.7

**Table 4. Information sources currently used by subjects for finding routes (Task 1), places (Task 2), properties of places and blocks (Task 3), and bikeability (Task 4). The table shows the number of subjects who mentioned each source and the mean usefulness for each source on a 5-point scale.**

not as useful as personal experience, this method is less time-consuming and has a wider reach. Also, as we suggest, a system that matches cyclists with similar opinions is likely to improve the utility of word of mouth.

The results of this section strongly support the utility of a geowiki for cyclists. Cyclists have much knowledge to enter, they judge the knowledge contributed by others to be useful, and this knowledge is not readily available elsewhere. We next consider the usefulness of our system as a *computational geowiki*.

## R2. Computation is more effective with user input

Our prior work has shown that cyclists need cycling-specific route-finding – that is, Google Maps, MapQuest, and their peers do not meet their needs [27]. Here we explore whether route-finding for cyclists is improved with user input.

Subjects entered 828 bikeability ratings on 637 blocks. Table 5 summarizes the effectiveness of three predictive methods, two generic and one based on user input. *Naive* is our simple bikeability formula that could be applied to all blocks. *CBF* is a formula published by the Chicago Bicycling Federation [3]; it could be applied to only 542 of the 637 user-rated blocks due to missing block attributes. *UAvg* predicts a user’s rating for a block as the mean of the other user ratings for that block (i.e., leave-one-out); this could be applied to the 136 blocks with at least two user ratings.<sup>15</sup>

The results were surprising. First, the two generic methods performed similarly, even though one (CBF) is a “serious” bikeability metric while the other (Naive) is an ad-hoc metric created to fill in the gaps where data to compute CBF were lacking. Second, the simple UAvg method modestly outperformed both generic methods.

These performance numbers are similar to good rating predictors in other domains when compared on MAE [19]. However, we conjecture that for transportation domains like

<sup>15</sup>In the user interface, estimated ratings use CBF if it is available and Naive otherwise.

	Naive	CBF	UAvg
Total user ratings	828	671	327
Mean absolute error (MAE)	0.80	0.74	0.65
Predictions in error by $\geq 1$ star	58%	54%	45%
Predictions in error by $\geq 2$ stars	16%	17%	11%
Predictable blocks	637	542	136
Blocks w/ mean error $\geq 1$ star	61%	52%	46%
Blocks w/ mean error $\geq 2$ stars	17%	18%	12%

**Table 5. Effectiveness of different techniques in predicting user bikeability ratings. Mean absolute error is the mean error of predictions; e.g., an average (user rating, CBF prediction) pair differed by 0.74 stars on a 5-star scale. Predictable blocks is the number of blocks where the technique could be applied.**

cycling, this level of error may still be too high. A typical prediction by all three methods was wrong by roughly one star or more, which corresponds (e.g.) to the difference between ratings of “fair” and “good” – semantically critical distinctions. Also, the cost of a poor route is high relative to other domains: while a moviegoer who dislikes a recommended movie can simply stop watching, a cyclist suddenly encountering bad parts of a recommended route may be stranded in an unfamiliar neighborhood.

In summary, simple user-driven methods may predict user bikeability ratings better than generic methods. Additionally, more sophisticated (and presumably better) generic methods will be hard to implement. They require even more data for each block – data often unavailable and difficult to collect even using a wiki, which excels at qualitative data but not the highly structured and quantitative data required. On the other hand, more sophisticated user-based methods such as collaborative filtering [21] require only block ratings, which our experiment shows are easy for users to provide. Therefore, the computational geowiki approach seems promising for the cycling domain.

## CONCLUSION: SUMMARY AND FUTURE WORK

We have produced a computational geowiki, a new type of system which enables interesting, useful, and novel geographic applications, overcoming significant interactivity and performance challenges to do so. Our empirical results in the context of a representative community, bicyclists, show that geowikis and computational geowikis are more effective than previous information gathering, exchange, and analysis tools. These results support the general promise of the technologies.

What are the implications of these results? First, route-finding should consider user-contributed information both when computing and presenting routes: route-finding algorithms work better when user input is considered, and route descriptions are enhanced by including user data (e.g., places along the way and comments about route segments themselves). Second, information resources should support off-topic conversation, rather than shunning it as noise, because it can enhance the communities surrounding them. Finally, wikis and other open-content systems should allow users to contribute opinions as well as facts.

We will deploy our geowiki, Cyclopath<sup>16</sup>, for public use by local cyclists in late summer 2008. From the perspective of cyclists, it will be a valuable resource (22 of 29 subjects asked when it would be available). From our perspective, it will comprise a valuable research platform, enabling us to explore important research questions, such as: How can we integrate the use of mobile location-aware devices? (Cyclists both need and acquire information while mobile.) What are the privacy implications of geowikis, and how can risks be mitigated? (1/3 of subjects expressed general concern about privacy when asked.) And how can user input be most effectively harnessed for automatic route-finding? Personalized algorithms like collaborative filtering seem very promising but suffer from severe sparsity problems. We plan to explore hybrid techniques, using collaborative-filtering-type approaches when possible and falling back to (or mixing in) simple averaging or generic methods when necessary.

#### ACKNOWLEDGEMENTS

Michael Ludwig wrote a substantial minority of our application's code, Andrew Sheppard helped categorize places and comments, and other members of our research group provided advice and support. Daan Hoogland translated Fiet-routeplanner from Dutch. This work was supported in part by the National Science Foundation, grant IIS 05-34692.

#### REFERENCES

1. Adler, B. and Alfaro, L. A content-driven reputation system for the Wikipedia. In *Proc. WWW*. 2007.
2. Balram, S. and Dragičević, S., eds. *Collaborative Geographic Information Systems*. Idea Group Publishing, 2006.
3. Barsotti, E. and Kilgore, G. The road network is the bicycle network: Bicycle suitability measures for roadways and sidepaths. In *Proc. Transport Chicago*. 2001. <http://bikelib.org/roads/roadnet.htm>.
4. Bearden, M. J. The National Map Corps (2007). [http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Bearden\\_paper.pdf](http://www.ncgia.ucsb.edu/projects/vgi/docs/position/Bearden_paper.pdf).
5. Bederson, B. B. et al. Enhancing in-car navigation systems with personal experience. In *Proc. Transportation Research Board*. 2008, 1–11.
6. Beenen, G. et al. Using social psychology to motivate contributions to online communities. In *Proc. CSCW*. 2004.
7. Bird Studies Canada. Christmas bird count (2008). <http://www.bsc-eoc.org/national/cbcmain.html>.
8. Bryant, S. L. et al. Becoming Wikipedian. In *Proc. GROUP*. 2005, 1–10.
9. Burrell, J. and Gay, G. E-graffiti: Evaluating real-world use of a context-aware system. *Interacting with Computers*, 14 (2002), 301–312.
10. Cosley, D. How oversight improves member-maintained communities. In *Proc. CHI*. 2005.
11. Cosley, D. et al. Using intelligent task routing and contribution review to help communities build artifacts of lasting value. In *Proc. CHI*. 2006, 1037–1046.

12. Cosley, D. et al. SuggestBot: Using intelligent task routing to help people find work in Wikipedia. In *Proc. IUI*. 2007, 32–41.
13. Counts, S. and Smith, M. Where were we: Communities for Sharing Space-Time Trails. In *Proc. ACMGIS*. 2007.
14. Espinoza, F. et al. GeoNotes: Social and navigational aspects of location-based information systems. In *Proc. Ubicomp*. 2001, 2–17.
15. Giles, J. Internet encyclopedias go head to head. *Nature*, 438 (2005), 900–901.
16. Goodchild, M. F. Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69.
17. Harper, F. M. et al. Predictors of answer quality in online Q&A sites. In *Proc. CHI*. 2008.
18. Hart, P. E. et al. A formal basis for the heuristic determination of minimum cost paths. *TOSSC*, 4, 2 (1968), 100–107.
19. Im, I. and Hars, A. Does a one-size recommendation system fit all? *TOIS*, 26, 1.
20. Kittur, A. et al. Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *Proc. alt.CHI*. 2007.
21. Konstan, J. A. GroupLens: Applying collaborative filtering to Usenet news. *CACM*, 40, 3 (1997), 77–87.
22. Lampe, C. and Resnick, P. Slash (dot) and burn: Distributed moderation in a large online conversation space. In *Proc. CHI*. 2004.
23. Lo, C. P. and Yeung, A. K. W. *Concepts and Techniques of Geographic Information Systems*. Prentice Hall, 2006, 2nd edition.
24. MacEachren, A. M. Cartography and GIS: Facilitating collaboration. *Progress in Human Geography*, 24, 3 (2000), 445–456.
25. NAVTEQ, Inc. NAVTEQ map reporter (2007). <http://mapreporter.navteq.com/>.
26. Priedhorsky, R. et al. Creating, destroying and restoring value in Wikipedia. In *Proc. GROUP*. 2007, 259–268.
27. Priedhorsky, R. et al. How a personalized geowiki can help bicyclists share information more effectively. In *Proc. WikiSym*. 2007, 93–98.
28. Ren, Y. et al. Applying common identity and bond theory to design of online communities. *Organization Studies*, 28, 3 (2007), 377–408.
29. Schafer, W. A. et al. Designing the next generation of distributed, geocollaborative tools. *Cartography and Geographic Information Science*, 52, 2 (2005), 81–100.
30. Sen, S. et al. tagging, communities, vocabulary, evolution. In *Proc. CSCW*. 2006, 181–190.
31. TomTom. Get to know TomTom MapShare (2007). <http://www.clubtomtom.com/general/get-to-know-tomtom-mapshare%E2%84%A2/>.
32. Viégas, F. B. et al. Studying cooperation and conflict between authors with history flow visualizations. In *Proc. CHI*. 2004.

<sup>16</sup><http://cyclopath.org>