bumpy, caution with merging: An Exploration of Tagging in a Geowiki

Fernando Torre, S. Andrew Sheppard, Reid Priedhorsky, Loren Terveen
GroupLens Research
Department of Computer Science and Engineering
University of Minnesota
Minneapolis, Minnesota, USA
{torre,sheppard,reid,terveen}@cs.umn.edu

ABSTRACT

We introduced tags into the Cyclopath geographic wiki for bicyclists. To promote the creation of useful tags, we made tags wiki objects, giving ownership of tag applications to the community, not to individuals. We also introduced a novel interface that lets users fine-tune their routing preferences with tags. We analyzed the Cyclopath tagging vocabulary, the relationship of tags to existing annotation techniques (notes and ratings), and the roles users take on with respect to tagging, notes, and ratings. Our findings are: two distinct tagging vocabularies have emerged, one around each of the two main types of geographic objects in Cyclopath; tags and notes have overlapping content but serve distinct purposes; users employ both ratings and tags to express their route-finding preferences, and use of the two techniques is moderately correlated; and users are highly specialized in their use of tags and notes. These findings suggest new design opportunities, including semi-automated methods to infer new annotations in a geographic context.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces; H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces—*Collaborative computing, computer-supported cooperative work, web-based interaction*

General Terms

Human Factors

Keywords

tags, wikis, geowikis, online communities, bicycling

GROUP2010, Novermber 7–10, 2010, Sanibel Island, Florida, USA.

1. INTRODUCTION

Tagging has become a ubiquitous information management technique on the Web. Users can apply tags – short textual labels – to items of interest and use these tags to browse and search for items. The design of tagging systems and the analysis of how tagging vocabularies evolve are active research areas [2, 4, 5, 11, 12, 15].

We report on a study of tagging in Cyclopath[1] [8, 10], a geographic wiki that provides route-finding services for bicyclists in the metropolitan area of Minneapolis-St. Paul, Minnesota, USA, an area of roughly 8,000 square kilometers and 2.3 million people. This is a new and distinctive context for the design and analysis of tagging systems for several reasons:

- **Geographic.** Users tag geographic objects in an interactive map. There are two distinct domains of geographic objects: *blocks*, atomic segments of the roads and trails that make up the transportation network, and *points*, places added by users that serve as the start and end points of routes and aid navigation.
- **Rich annotation ecosystem.** We introduced tags into Cyclopath about nine months after the site went live. Two other annotation techniques had been available from the beginning: *notes*, free-form textual comments attached to points or blocks, and *ratings*, subjective opinions of the bikeability of blocks.
- **Novel application.** We modified the Cyclopath route finder to let users express routing preferences with tags and thus fine-tune the routes they received.

We pose three research questions to understand the tagging behaviors that have emerged in Cyclopath:

RQ1. Vocabulary. *How do the tagging vocabularies that evolve for the two object domains – blocks and points – compare? What factors drive the evolution of tagging vocabulary in the Cyclopath domain?* Looking at the vocabulary in a tagging system is one of the most common ways to gain an understanding of how tags are used in the system. We found that the two object domains in Cyclopath developed almost wholly distinct vocabularies with quite different characteristics; we argue that the differences emerge because the two types of objects support quite different tasks.

[1] http://cyclopath.org

RQ2: Ecosystem. *How do users employ the three annotation techniques tags, notes, and ratings? Do they express overlapping or distinct content?* Having three closely related techniques made it natural to investigate the interactions and relationships between them. We found significant content overlap between notes and tags; we computed the notion of tag-derived preference and found a positive correlation between tag-derived preferences and ratings.

RQ3: Specialization. *How do individuals balance annotation roles, techniques, and objects? Are they specialists or generalists?* The next step was to study tags (and annotations in general) at the user level. We found strong specialization: users tend to either use or apply tags, not both, and annotate either blocks or points with either notes or tags.

The rest of the paper is organized as follows. We first sketch related work and then describe the design of tagging in Cyclopath. We devote most of the paper to describing the methods we used to investigate our research questions, our findings, and their implications. We conclude with a brief summary and discussion of future work.

2. RELATED WORK

2.1 Geowikis and geotagging

Many web sites let end users edit geographic content. Google Maps lets users edit the locations of searched-for places and add new places. Google My Maps goes further, enabling collaborative editing of geographic points, paths, and polygons, all of which can be annotated with text, images, and videos. Sites like FixMyStreet and SeeClickFix let users plot the location of potholes and similar problems on a map. Open Street Map is a large project to build a worldwide street map with a wiki. Google Map Maker lets users directly edit Google Maps data (in some countries) and submit those changes for inclusion in the public map. There is a growing body of scholarly work on these systems, including multiple studies of Cyclopath [8, 9, 10] and an analysis of how FixMyStreet facilitates citizen-government interaction [6].

We extend previous work by examining the design choices required to add tagging to a geowiki and analyzing a year's worth of user tagging behavior.

Geotagging is a distinct idea: adding geographical metadata (typically latitude and longitude) to content like web pages and photographs (as in Flickr). We look at almost the opposite situation: adding tags to existing geographic data.

2.2 Tagging

Tagging is an active research area, including studies of algorithms to suggest tags to users [3, 12], evaluations of tag clouds [11], and studies of tagging in an enterprise information sharing system [7]. Two strands of work are most relevant to ours: analysis of the vocabularies that emerge in tagging systems and studies of how tags relate to associated techniques like recommender systems.

Vocabulary analysis. Research includes information theoretic analysis of the overall structure and evolution of tagging vocabularies [2], empirical analysis of what constitutes "quality" for tags [13], and several schemes for categorizing tags [4, 12]. We use the categories introduced by Sen et al. [12] to analyze the Cyclopath tagging vocabulary.

Tags and associated techniques. A large body of work explores various ways to integrate tags and recommender systems, e.g., using tags to infer user preferences for items and improve recommendation algorithms [14, 16, 21] and using tags to explain recommendations [18]. Other work has investigated the relationship between the text of web pages and the tags applied to web pages, finding that 50% of tags were present in the text of web pages to which they were applied [5]. In this research, we examine the relationship between preferences expressed through tags and through ratings, and we compare the content and usage patterns of tags to textual annotations.

2.3 Specialization

Prior research on a variety of online communities has found that users take on specialized roles. In online discussion forums, Turner et al. [17] and Welser et al. [19] identify different roles that users assume, notably "Question Person" and "Answer Person". In Wikipedia, Welser et al. [20] mapped out various roles that contributors can play, such as technical editors, substantive experts, vandal fighters, and social networkers. Bryant et al. [1] found that Wikipedia editors shifted concerns as they became more experienced, evolving from a focus on topics about which they had some personal expertise to taking on different types of "community maintenance", e.g. monitoring for vandalism and enforcing policies like "Neutral Point of View".

Compared to prior work, we investigate a particular type of specialization: how Cyclopath users balance their choice of annotation roles, techniques, and objects.

3. THE CYCLOPATH TAGGING SYSTEM

Cyclopath is a geographic wiki and route finding system for bicyclists which has been publicly available since August 2008. Over 2,000 people have registered for the site. During riding season, several dozen users log in each day, more than one hundred anonymous users visit, and collectively they request 150-250 routes per day. Users have rated about 70,000 blocks and made over 10,000 revisions to the map.

3.1 Design Choice: Wiki-tags

We faced a number of choices when we designed the Cyclopath tagging system. Most notably, we had to decide whether tags should be normal objects in the wiki model. For our purposes, the key properties of wiki objects are:

- They are *public*. All objects may be changed or deleted by any user. Objects do not have owners.
- Each set of changes (and deletions) results in a new *revision* that captures the changes. There is only one current version of an object, but previous versions are saved. Any user may revert (undo) any revision, rolling back to a previous version.

We can interpret the implications of making tags normal wiki objects in the framework of Sen et al. [12]. Wiki-tags would be completely open or public on the *sharing* dimension. For example, if the tag *construction* is applied to a block, this is visible to all users. Wiki-tags are owned by the community, and there is a single set of tags for any object; thus wiki-tags have *broad scope*. For example, the tag *construction* can be applied at most once to a given block.

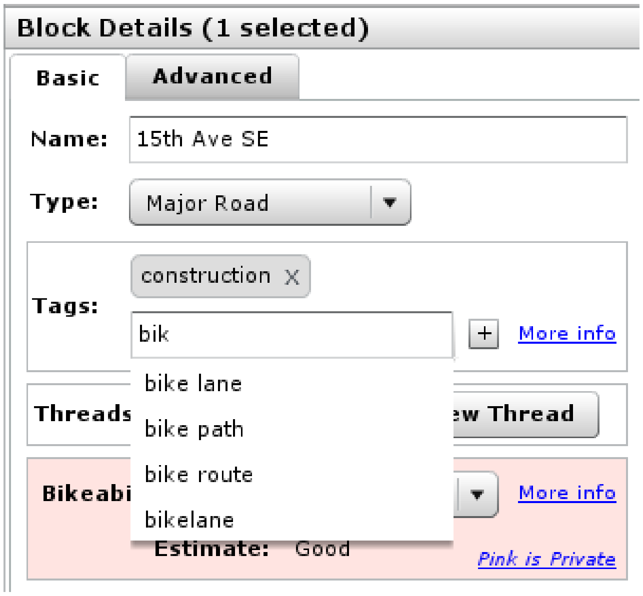

Figure 1: Tag application interface. The tag *construction* has already been applied to this block. The user is entering a new tag (so far "bik"), and auto-complete shows existing tags that begin with the text entered so far.

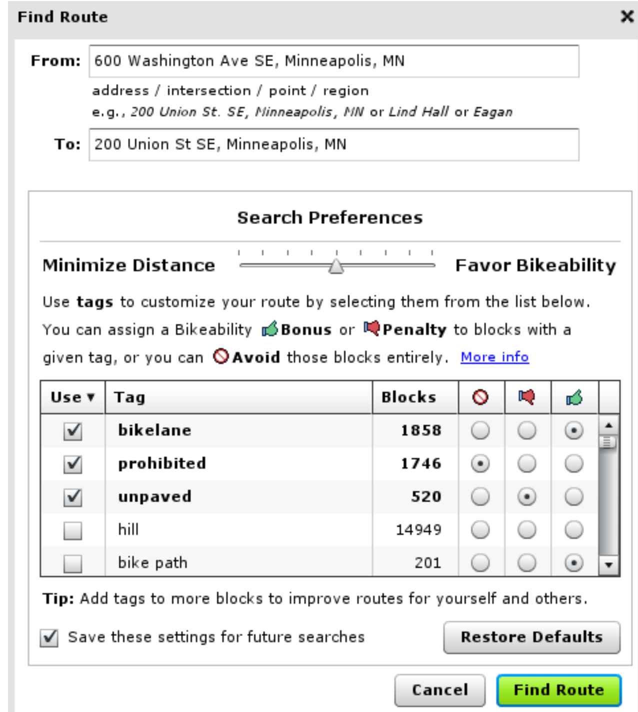

Figure 2: Route finder and tag preferences interface. In this example: (a) blocks with the *bikelane* tag are favored, (b) blocks with *unpaved* are penalized, (c) blocks with *prohibited* are avoided completely, and (d) all other tags (in particular, *bike path* and *closed*) have no effect.

We could have implemented tags as a separate add-on to the wiki or by extending the wiki model (e.g., by remembering the user who applied a tag to an object, and making that tag application only editable by that user). However, we prefer to follow the wiki model whenever possible. Moreover, we wanted to promote the use of tags that were useful to the community. Therefore, implementing tags as wiki objects was a good fit for our goals.

3.2 Entering and using tags

Tags are applied to points or blocks using a rather standard interface (see Figure 1). Users type into a text field, and an auto-complete function suggests tags that have been applied to this type of object. This encourages users to reuse tags.

In addition to enhancing user exploration, tags are used in two system functions. (a) Point tags can be used to filter the display of points on the map. For example, a user could request to see only points that had one of the tags *air pump* or *bike rack*. Filtering is rather straightforward, and we do not discuss it further in this paper. (b) Block tags can be used to tailor route-finding preferences.

Route-finding preferences. Before we introduced tags, our route-finding algorithm estimated the bikeability of a block by considering user ratings if available, and by using objective features (such as block type, speed limit, and shoulder width) otherwise. We believed that tags would be an effective complement to ratings. A user can have only one rating for a given block – one of five choices from "Impassable" to "Excellent" – but this does not capture why a user likes or dislikes a block. Hence, we modified the route-finding interface to let users express preferences for or against certain tags (see Figure 2). Users may assign selected tags a bonus or a penalty or state that they be avoided altogether. (User tag preferences are saved between route requests.) This innovative use of tags gives users more expressivity and control. For example, a user might assign a penalty to the *hill* tag in most requests, but assign this tag a bonus when in the mood for a challenging workout. The route finding algorithm respects any specified tag preferences. If a block has any tag that the user wants to avoid, that block will not be added to the route. If a block has a tag that has been assigned a bonus (or penalty), the block's bikeability estimate will be incremented (or decremented).

3.3 Annotation framework

Tags have both similarities and differences with the two other Cyclopath annotation techniques, notes and ratings. Like notes, tags can be applied to blocks or points, are public, and aid users in evaluating points, blocks, and routes. However, like ratings, tags are used by the route finding algorithm. Table 1 summarizes the three annotation techniques.

4. TAGGING USE AND DATA

Tagging was added to Cyclopath in April 2009. In the subsequent year, 178 users have applied tags, creating a vocabulary of 239 distinct tags that have been applied over 1,900 times to over 1,400 distinct blocks and points. (There would be more tag applications if tags had *narrow* (multiple tag applications per user for each object) rather than *broad* (a single set of tag applications for each object) scope [12]). Over one quarter of all revisions include tag edits. Over 600 users have filtered with tags and over 2,000 users have ex-

Annot.	What	Who	Used by
ratings	blocks	private	**machine** route finding algorithm
notes	blocks, points	public	**human** evaluate blocks and points
tags	blocks, points	public	**human and machine** filter points express route preferences evaluate points and blocks route finding algorithm

Table 1: Framework for Cyclopath annotations.

	Blocks	Points
tags	91	163
tags (non-auto-applied)	86	163
tag applications	20,030	973
applications of non-auto tags	940	973
objects with tags	19,525	668
objects with non-auto tags	862	668
tag applications removed	626	12

revisions to Cyclopath since tags went live	3406
revisions involving tags	919
users that have applied tags	178
users that have removed tag applications	49
users that have filtered using tags	623
users that have used tag routing preferences	2089

Table 2: Usage statistics.

pressed route finding preferences with tags. As common in open content systems, many more users consume information (here, using tags for filtering and route finding) than produce information (here, applying tags).

The tagging vocabulary contains a number of tags that were applied automatically. Prior to the introduction of tagging, blocks could have any of three binary features: *unpaved*, *bike lane*, and *closed*. We automatically converted these features into tags, thus creating a small initial tag vocabulary consisting of 2,470 applications of these three tags. Later in 2009 we wrote code to automatically apply two additional tags. *prohibited* was applied to about 1,700 blocks (mainly expressways) where it is illegal to ride a bicycle. *hill* was applied to almost 15,000 blocks with a slope greater than 2%, based on existing measures of hilliness and existing applications of the tag in the system. We also set default route finding preferences to avoid both the *closed* and *prohibited* tags. We refer to this set of five tags as auto-applied tags.

In our analysis of tag applications, we ignore all auto-applied tags and any tags created by members of the Cyclopath research team. In our analysis of tag preferences, we ignore preferences for *closed* and *prohibited*, because these tags have default preferences in the route finding dialog.

Table 2 summarizes the tagging data that we analyzed. In the next three sections, we sketch our methods (a combination of manual coding and quantitative analysis), present our findings, and discuss their implications.

5. RQ 1. VOCABULARY

In this section, we analyze the tagging vocabulary that has evolved in Cyclopath and identify factors that influenced its development. A *tagging vocabulary* is the set of distinct tags used in an online community. For example, if the tag *low traffic* has been applied twice, *rocky* five times, and *gravel* once, then the vocabulary would be the set {*low traffic*, *rocky*, *gravel*}.

As mentioned earlier, blocks and points are intrinsically different and play different roles in Cyclopath: blocks are the components of routes, and points are landmarks and the start and end points for routes. Further, block tags and point tags serve different purposes: block tags can fine-tune route preferences, while point tags filter the map. Thus, we provisionally treat block tags and point tags as two separate vocabularies and do several analyses to see if this separation is warranted.

5.1 Tag diversity

The first characteristic that we examine for the two tagging vocabularies is diversity, the ratio of the number of distinct tags to the number of tag applications. High diversity means that each distinct tag has been applied just a few times; low diversity means that each distinct tag has been applied many times. Cyclopath point tags have a diversity of 0.18, and block tags have a diversity of 0.09. Therefore, the mean point tag has been applied about 5 times, and the mean block tag about 10 times. However, the mean number of applications is a misleading statistic: both vocabularies follow the expected non-normal "long tail" distribution. Indeed, the reason the point tag vocabulary is more diverse is because it has a longer "tail", i.e., tags that have been applied just once.

Contributing to the lower diversity of block tags are the inherent nature of blocks and a specific feature of the Cyclopath user interface. Contiguous sequences of blocks often share a feature – for example, several blocks in a row may be bumpy or scenic. Users can select a set of blocks, then apply the same tag to the entire set with a single UI action. On the other hand, points must be evaluated separately, and point tags applied one by one.

A difference in diversity tells us about the global structure of the two tagging vocabularies. However, it does not tell us about similarities or differences in their content. We examine this issue next.

5.2 Overlap

A natural way to compare the two tag vocabularies is simply to compute the overlap (intersection). As background, we note that the auto-complete feature (see Figure 1) could bias user tag application decisions. To avoid biasing users in favor of developing separate tag vocabularies for points and blocks, we initially designed auto-complete to suggest tags applied to both blocks and points. However, after about six months of use, only 7 out of 109 tags occurred in both vocabularies. To better support emergent practice, we then modified auto-complete to suggest only tags from the relevant vocabulary for each object. The total overlap is now 10 tags out of 239 distinct tags.

For those tags that appear in both vocabularies, the majority of their applications are in just one. For example, the tag *steep hill* has been applied to 20 blocks, but just one point. The tag *food* has been applied to 67 points, but

	Blocks	Points
factual	58%	91%
subjective	35%	5%
personal	1%	0
other	2%	2%
no agreement	3%	2%

(a) factual/subjective/personal

	Blocks	Points
bike-specific	70%	6%
non-bike-specific	21%	91%
other	1%	1%
no agreement	7%	2%

(b) cycling content

	Blocks	Points
noun (is-a)	33%	49%
noun (has-a)	20%	37%
adjective	34%	5%
verb	2%	4%
other	3%	2%
no agreement	8%	3%

(c) parts of speech

Table 3: Results for content categorizations.

to just two blocks (which perhaps have many restaurants alongside them).

Since blocks and points are different types of objects, it is not surprising that different terms are used to describe them. However, perhaps a more abstract categorization of the two vocabularies would reveal deeper similarities – or clarify their differences. We did a content categorization to investigate these possibilities.

5.3 Content Categorization

We did three categorizations of the tagging vocabularies:

- **Factual, subjective, personal.** Sen et al. [12] used these categories to analyze the MovieLens tagging vocabulary. They also found that different categories were useful for different user tasks.
- **Cycling content.** Priedhorsky et al. [10] categorized Cyclopath notes, with a fundamental distinction between notes that were and were not directly related to cycling. They found that while most point notes were not related directly to cycling, most block notes were.
- **Part of speech.** A very general way of distinguishing vocabularies is in terms of the parts of speech of the individual tags.

For each content categorization, three coders were given classification rules and categorized the tags independently. Tags that did not fit into any of the categories were classified as *other*. After independent classification, the tags were assigned their final categories using majority rule.

5.3.1 Factual, subjective, personal categorization

Building on the definitions by Sen et al., we define factual, subjective, and personal tags as follows:

1. **Factual** tags express objective properties of an item. For example, in Cyclopath, *bridge* and *one way* are factual block tags, and *bike shop* and *pizza* are factual point tags.
2. **Subjective** tags express opinions. This class can be quite broad, since even if nearly everyone would agree on the application of a tag, we classified it as subjective if disagreement was reasonable. Some examples are *scenic* and *dangerous* for blocks and *bad coffee* and *awesome* for points.
3. **Personal** tags have the user who applied them as the only intended audience. An example of a personal tag in Cyclopath is *home*.

The results of this categorization are shown in Table 3(a). All three coders agreed 78% of the time for blocks and 90% of the time for points, while at least two of the coders agreed 97% of the time for blocks and 98% of the time for points. In comparison, the distribution of tags in MovieLens was 63% factual, 29% subjective, 3% personal, and 5% "other" [12].

There were few personal tags in either vocabulary; indeed, most personal tags that were applied were removed by other users. One such example is the tag *work1*, for which 74 applications to blocks were removed.

We expected that community ownership of tags would lead to a very high proportion of factual tags. And factual tags indeed are the majority in both vocabularies. However, while over 90% of point tags are factual, the distribution of the vocabulary for block tags is more comparable to that of MovieLens (where tag applications are owned by users, not the community), and there is a sizeable minority of subjective block tags. We can think of two possible explanations for this. First, perhaps Cyclopath users in general simply agree that a particular block is *bumpy* or *dangerous*, even though it is theoretically possible to disagree. Second, this may indicate an intrinsic problem with the Subjective category or at least our application of it. Maybe "subjective" is subjective – or maybe we were too strict in our application.

Why do we see a different distribution for point tags? To answer this question, we refer back to Sen et al. They identified five user tasks supported by tags in MovieLens: self-expression, organizing, learning, finding, and decision support. Surveys showed that different classes of tags were more useful for different tasks. Factual tags were particularly useful for learning and finding, while subjective tags were useful for self-expression and decision support. The learning and decision support tasks are most important in Cyclopath: learning is supported by map exploration, and decision support occurs during route finding and evaluation. Point tags are not used during route finding. Therefore, they are mainly used for learning, which explains the high percentage of factual tags. On the other hand, block tags must support both learning and decision support, which may account for the higher proportion of subjective tags.

5.3.2 Cycling content categorization

We also distinguished tags by whether they are bike-specific:

1. **Bike-specific** tags are directly related to the practice of bicycling. Block tags like *caution with merging* and *moderate traffic* and point tags like *free air* and *covered bike parking* fall into this category.

2. **Not bike-specific** tags are everything else. These include block tags like *river* and *storefronts* and point tags like *church* and *fast food*.

Table 3(b) shows the results for this categorization. All three coders agreed 40% of the time for blocks and 90% of the time for points, while at least two coders agreed 93% of the time for blocks and 98% of the time for points.

As with notes [10], a much higher proportion of block tags are bike-specific. This is because when planning and evaluating routes, cyclists are concerned with properties of the route (or the blocks that comprise it) that affect bikeability. On the other hand, points are of more general interest as possible destinations, landmarks, or resources along the way; therefore, cyclists likely are more interested in intrinsic information about the places themselves, e.g., is it a good place to get food or an interesting place to visit?

5.3.3 Parts of speech categorization

We used the following rules to categorize by part of speech:

1. **"Is-a" noun** tags fit the pattern "this block/point is a(n) X". Examples include *alley* and *highway* for blocks and *bakery* and *gas station* for points.

2. **"Has-a" noun** tags fit the pattern "this block/point has X". Examples include *curb* and *no shoulder* for blocks and *internet* and *coffee* for points.

3. **Adjective** tags describe a property of a block or point. Examples include block tags like *quiet* and *rough* and point tags like *expensive* and *awesome*.

4. **Verb** tags describe actions relevant to a point or block, such as the block tag *avoid* and the point tag *get the fudge cake*.

The results for this categorization are shown in Table 3(c). All three coders agreed 56% of the time for blocks and 62% of the time for points, while at least two of the coders agreed 92% of the time for blocks and 97% of the time for points.

Once again, the categorization yields quite different results for both vocabularies. The most striking difference is in the use of adjectives. About a third of all block tags are adjectives, while only 5% of tags applied to points are adjectives. These adjective block tags give information about what it's like to ride there. On the other hand, point tags often describe what a place is or what can be found there, uses well-supported by nouns.

5.4 Results Summary

Cyclopath is unusual in that two separate tagging vocabularies with very different characteristics have emerged. The block tag and point tag vocabularies differ not only in their specific terms, but also in the type of content and distribution. Three content categorizations revealed unique, yet mutually supporting points of comparison. Point tags tend to be factual nouns and not bike-specific, while block tags are a mix of factual and subjective nouns and adjectives and are mostly bike-specific. We trace these distinctions to intrinsic differences between points and blocks, the different roles they play in Cyclopath, the different roles played by block and point tags, and user interface differences.

6. RQ 2. ECOSYSTEM

Tags, notes, and ratings comprise an ecosystem of annotation techniques in Cyclopath. Table 1 (in Section 3 above) summarizes the important properties of each technique. Here we investigate similarities and differences in the usage patterns of tags vs. notes and tag preferences vs. ratings.

6.1 Tags vs. Notes

Tags and notes have obvious similarities. Both describe blocks and points, and both are public, owned by the Cyclopath community. As is typical, Cyclopath tags are mostly single words, with a few short phrases. On the other hand, notes can be arbitrarily long texts. For example, here is a Cyclopath block note: *Buses go very fast along here, but they generally give you plenty of space. No stop signs!*, and here is a point note: *They fix and give away commuter bikes, on volunteer time and donations. Good source of used parts.* We examined two relationships between tags and notes: content overlap and usage substitution.

Content overlap. We examined cases where both tags and notes had been applied to a single block. For example, the tag *dangerous* had been applied to a certain block, and that block also had the following note:

> *The bike lane on the left side abruptly disappears in Dinkytown and reappears after passing through w/ dangerous parking situation. It reappears only to disappear again at the ramps to 35W. Here it is totally unsafe...*

Thus, the tag *dangerous* appeared literally in the note.

Our overlap analysis was simple: for each block b with tags $t_1, ..., t_N, N > 0$ and notes $n_1, ..., n_M, M > 0$, check whether t_i appears in any of the notes $n_1, ..., n_M$, and the same for points. 31% of block tag applications and 26% of point tag applications were mentioned in a corresponding note.

Usage substitution. Given the similarity of tags and notes, we wondered how the introduction of tags 9 months after Cyclopath went live affected note usage. We thus calculated the mean number of note edits per revision before and after tags were introduced. There was a significant difference in the mean number of note edits per revision before and after tags were introduced: 0.87 vs. 0.66 (t-test; $p<0.01$). This suggests that tags assumed some of the functions notes had served.

6.2 Tag Preferences vs. Ratings

As we explained earlier, both ratings and tags play a role in the Cyclopath route finder, creating a context for comparing the two. We show how the tag preferences users express when requesting routes can be used to derive user preferences for blocks, then investigate whether these tag-based block preferences correlate with the preferences users express through ratings.

6.2.1 Tag Preference/Ratings Correlation

Ratings let users express specific bikeability preferences for specific blocks: *I think this block of Summit Avenue is "excellent"* or *I think this block of Hennepin Avenue is "poor"*. Tags let users express generic preferences for the types of blocks they want in a route: *I prefer blocks with the tag "bike-lane" (a bikeability "bonus") and dislike blocks with the tag*

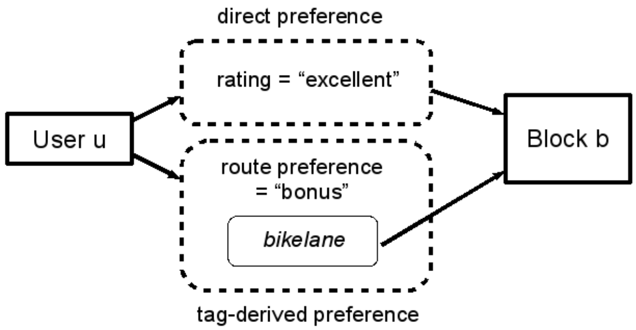

Figure 3: The relationship between ratings (directly expressed preferences) and tag-derived preferences.

"bumpy" (a "penalty")". Consider the following situation, as illustrated in figure 3: a user u has rated a block b as "excellent"; the tag *bikelane* has been applied to b (not necessarily by u); u has expressed a "bonus" route-finding preference for the tag *bikelane*. u's (directly expressed) ratings-based preference for b and u's (derived) tag-based preference for b appear consistent. This makes sense: if a user likes blocks with *bikelane*s, shouldn't the user rate block b highly?

Our goal is to test this intuition formally. We first introduce some terminology: a user has a *tag-derived preference* for a block b if there exists at least one tag t such that (a) t has been applied to b, and (b) the user has expressed a route-finding preference using t. We then compute the global correlation between ratings and tag-derived preferences by aggregating across all users, blocks, and tags. We next show how to compute *tag-derived preferences*.

6.2.2 Tag-derived preferences

The intuition for computing tag-derived preferences is simple. Suppose the tags t_i, and t_j have been applied to a block b. If a user u expressed a "bonus" preference for t_i, this should lead to a positive tag-derived preference for b, and if the user expressed a "penalty" or "avoid" preference for t_i, this should lead to a negative tag-derived preference for b. And if the user expressed opposing preferences for t_i and t_j, i.e., "bonus" for one and "penalty" or "avoid" for the other, this should lead to a more-or-less neutral tag-derived preference for b. To formalize these intuitions, we must make two decisions: (a) how to translate tag preferences to numeric values,[2] and (b) what to count as a user having a preference for a tag. We consider each of these in turn.

Translating tag preferences to numeric values. We must specify precisely how to go from *bonus*, *penalty*, and *avoid* to numeric values. It is easy to specify constraints on possible values. (a) Values should range from -1 to 1. (b) A mathematical relationship that must hold among the three preferences is: $-1 \leq$ avoid $<$ penalty $< 0 <$ bonus ≤ 1. (c) Since the "avoid" preference is as negative as possible, it makes sense to set its value to -1. (d) "Penalty" and "bonus" are opposites (conceptually and in the Cyclopath route finder), so it makes sense for "penalty" to be equal to "$-$bonus. However, there are no obviously correct values for "penalty" and "bonus". Therefore, we experimented with a number of different values, including 0.75, 0.5, and 0.1. When we computed correlations between tag-derived preferences and ratings using the different values, the results were virtually identical unless we gave "penalty" and "bonus" scores close to 0. When we present the correlations below, we present them using two sets of values for ("avoid", "penalty", "bonus"): (-1, -0.75, 0.75) and (-1, -0.1, 0.1).

[2]Recall that ratings already are on a numeric scale of 0 ("impassable") to 4 ("excellent").

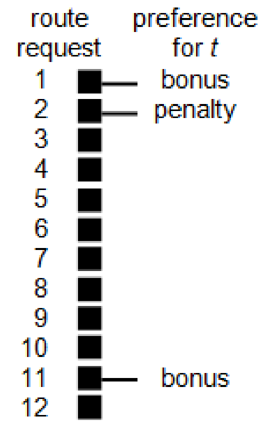

Figure 4: Example use of tag preferences in route requests used for illustrating three ways to count tag preferences.

Tag preferences: what to count. What it means for a user to have a preference for a tag turns out to be a bit complicated, since users may change their preferences with each route request (recall that preferences are saved between requests). Figure 4 illustrates this situation. A user has issued a total of 12 route requests, expressing a "bonus" preference for the tag t on the first request, changing this to a "penalty" on the second request, then returning to a "bonus" on the eleventh request. t could be a tag like *hill*, and this pattern could reflect alternation between a weekend preference (*bonus hills for a better workout*) and a weekday preference (*avoid hills for a less sweaty commute*). There are three plausible things we could count to compute tag preferences:

1. **End-state.** Define a user's preference for a tag as the most recent preference expressed for the tag. The intuition is that users may experiment for a while, but have one true preference that they eventually reach. In figure 4, only "bonus" would be used to compute the preference for tag t.

2. **Decision-average.** Define a user's preference for a tag by weighting equally each decision to change a preference for the tag. (This and the following definition do not assume that a user has one true preference). In figure 4 there are three decisions regarding t: two to "bonus" it and one to penalize it. Thus, "bonus" would be weighted twice as much as "penalty" in computing the preference for t.

3. **Use-average.** Define a user's preference for a tag by weighting how many times it was used in a route request. In figure 4, "bonus" was used in three route requests and and "penalty" in nine requests. Thus, "penalty" would be weighted three times as much as "bonus" in computing the preference for t.

	end-state	decision-average	use-average
(-1,-0.75,0.75)	0.259	0.278	0.283
(-1,-0.1,0.1)	0.149	0.211	0.225

Table 4: Correlation between tag-derived preferences and ratings. Two different numeric translations of the tag preferences (avoid, penalty, bonus) are used.

6.2.3 Results and Analysis

The results of the correlation between tag-derived preferences and ratings are given in Table 4. The correlation is positive, but weak to moderate (results were similar for all numeric translations of tag preferences we used). However, the average correlation obscures individual cases where the correlation is particularly good or bad. Considering such cases is instructive in understanding the relationship between tag-derived preferences and ratings.

Good correlation. There is a block of the Minnehaha Creek Trail that includes the tags *stairs* and *pedestrian bridge*. One user gave the tags *stairs* and *pedestrian bridge* penalties and also rated the block as *poor*. Another user also gave *stairs* a penalty and rated the block as *poor*. In another case, a block along the Kenilworth Trail with the tags *bikelane* and *bike path* was rated as *excellent* by six users who also gave these two tags a bonus. It is probable that the qualities represented by the tags in these two examples influenced the user's preferences for the blocks.

Poor correlation. There is a block of Dale St. that has the *bikelane* tag. From the pool of users who rated that block, only *bonus* was assigned to *bikelane*. But most users rated the block as *poor* or *fair*, while nobody rated it as *excellent*, resulting in a strong negative correlation between the tag-derived preferences and the ratings. A possible reason for this can be inferred by looking at a note attached to this block: *Surface of road is very bumpy and full of pot holes. Watch where your tires go!*. If the tags *bumpy* or *pot holes* had been added to this block, the tag preferences might have reflected the user ratings much better! Therefore, this correlation failure seems to be caused by missing data. A second reason for poor or negative correlation may be implicit priorities among tags: for example, a user might normally avoid *unpaved* bike paths, but will make an exception if they are *scenic*. Finally, some tags like *construction* reflect time-limited properties of a block, but we do not see evidence of users changing their ratings to reflect such temporary changes in the bikeability of a block.

6.3 Results Summary

We showed that tags and notes have significant overlap in content. Rather than seeing this as wasteful redundancy, we view it as confirmation that tags and notes have distinct utility: tags provide quick descriptions of objects and let users fine tune the route finder; notes provide detailed information useful for learning and evaluation. In addition, there was a positive correlation between user preferences expressed directly as ratings and derived from the use of tags in route requests. We further identified situations where the correlations were particularly positive or negative. We suggest follow-up research on both sets of results in Section 8.1.

Role	Count	
consumers	379	89%
producers	26	6%
non-specialists	21	5%

(a) tagging role

Technique	Count	
notes	52	39%
tags	39	29%
non-specialists	43	32%

(b) annotation technique

Technique	Object	Count	
notes	blocks	170	84%
	points	14	7%
	non-specialists	18	9%
tags	blocks	36	50%
	points	23	32%
	non-specialists	13	18%
notes & tags	blocks	183	77%
	points	32	13%
	non-specialists	23	10%

(c) annotation object

Table 5: Specialization in annotation behavior.

7. RQ 3. SPECIALIZATION

The findings for the previous research question illuminate the overall relationship among the use of tags, notes, and ratings. However, we also investigate similar issues from the perspective of users. Specifically, we examine how users balance their tagging role (producer vs. consumer), public annotation technique (tags vs. notes), and annotation object (blocks vs. points): are they specialists or generalists?

7.1 Tagging role: producer vs. consumer

Most users of open content systems only consume information, and only a small minority actually contribute it. We examined whether this was true for Cyclopath tagging. Specifically, we looked at tagging actions concerning blocks – either applying tags or using tags to express route finding preferences.[3] We considered only users who made at least 5 tagging-relevant actions with blocks – any combination of applying tags and using tags to specify route finding preferences. We categorized users as specialists if at least 75% of their tagging-relevant actions was either production or consumption. Table 5(a) presents the results. As expected, the vast majority of users (89%) were information consumers. However, we were interested to see that the majority (55% or 26 of 47) of the remaining users were production specialists; they produced value while taking little value in return. This finding is related to work on social roles in online discussion forums that identifies "answer people", users who primarily answer rather than ask questions [19]. However, prior research did not have access to private user behavior such as message reading; thus, it is possible that "answer people" received value from reading answers even if they didn't pose many questions of their own. Because we have access to the entire Cyclopath usage history, we can analyze behavior like message reading or preference setting that is typically unavailable to researchers [8].

[3]We did not examine tagging actions concerning points because we did not log point filtering behavior.

7.2 Annotation technique: tags vs. notes

Tags and notes are both publicly visible, shared annotations. The previous section showed that both techniques are used, and that the introduction of tagging decreased the use of notes. But do individual users combine both techniques? Or do they specialize?

We compared tag applications and note edits made after tags were added to Cyclopath. We considered the set of users who had made at least five public annotations (tag applications or note edits) and defined users as specialists if at least 75% of their annotations were either tags or notes. The results are shown in Table 5(b). We make several observations. First, a majority – 68% – of users specialized in one technique, with note specialists outnumbering tag specialists 52 to 39. Second, a healthy proportion of users edited both notes and tags. We think that this may be because both types of edits are easy to make, and they serve complementary purposes, as we sketched in Table 1 in Section 3. Finally, we did followup analysis to investigate how the introduction of tags into Cyclopath affected users who had edited at least 5 notes prior to this time. We saw an utterly mixed picture. The vast majority of such users (99 out of 122) had become inactive by the time tags were introduced. Of the other 23 users, 8 remained note specialists, 8 became tag specialists, and 7 became generalists. Additional work is required to understand what drove individual decisions to retain an existing practice or adopt a new one.

7.3 Annotation object: blocks vs. points

Finally, we wondered whether users preferred to annotate one type of object, either blocks or points. As in the previous analysis, we considered users who had a total of at least 5 tag edits or note applications. We defined users as specialists if at least 75% of their annotation were associated with either points or blocks. We did this analysis three ways: with tags alone (i.e., only for users who applied at least 5 tags), with notes alone (only for users who edited at least 5 notes), and with tags and notes in combination (counting users whose total of note edits + tag applications was at least 5). Table 5(c) presents the results and lets us make several observations. First, all three analyses reveal that a high proportion of users are specialists. Second, there was a noticeable difference in the relative proportion of point and block specialists when we only considered notes and when we only considered tags. For notes, 92% of specialists (170 out of 184) specialized in blocks. For tags, the ratio is much more even: 61% of specialists (36 of 59) specialized in blocks.

7.4 Results Summary

We found strong specialization in usage of the public annotation techniques tags and notes. (a) Most users consume annotations; only a small minority produce them. This result is consistent with and extends prior research on open content systems. (b) Most users who do annotate favor one technique, with notes more popular. (c) Users also specialize in their choice of objects to annotate, and most of those who specialize are block specialists. We speculate that this is because blocks play a central role in Cyclopath (they are the building blocks for routes), and there are many more blocks than points (over 150,000 vs. about 2,400).

8. CONCLUSION

8.1 Implications for Design and Research

This work raises several interesting followup possibilities.

First, there are more opportunities to analyze the ecosystem of Cyclopath annotation techniques, notably deeper exploration of users' route preferences and how they express them. User changes to their tag preferences and ratings form an intriguing context for studying these questions. We hypothesize several different types of preference changes:

1. A more-or-less **permanent change** in preference: for example, as cyclists gain experience, they may be willing to ride on roads they previously considered too busy or dangerous.
2. A **temporary change** in preference: for example, a cyclist might avoid bumpy and unpaved bike paths in general, but will seek them out when trying out a new mountain bike.
3. A **change in external conditions**: for example, a traffic detour might turn a previously quiet and easily bikeable road into a dangerous adventure.
4. **Experimentation**: for example, a user might request a route, dislike the results, and then (re-)rate some blocks or modify some tag preferences to try to get a better route.

We conjecture that certain usage patterns indicate different types of preference changes. Do users request a route, change a preference, then immediately re-request the same route? If so, this may indicate experimentation. (We already have found that 48% of all tag preference changes occur during route re-requests.) Does a user maintain a preference for a long time, then change it and keep the new one for a long time? If so, this may indicate a true change in preference. Does a user have a preference, change it for a route request, then change it back for the next request (of a different route)? This may indicate a temporary change in preference. Appropriate methods to investigate these issues include both quantitative analysis of logged usage data (as we did in this paper) and user interviews.

Second, our findings suggest a number of semi-automated methods to infer new annotations.

1. **Notes to tags.** An obvious way to infer new tags is to mine the text of notes. The easiest case is when an existing tag is contained within a note. For example, one note in Cyclopath reads: *A nice scenic stretch to include on any route. Very nice shoulders and traffic isn't heavy.* An already existing tag is *scenic*. Others include *low traffic* and *moderate traffic*. Slightly more sophisticated techniques could be used to infer these tags. Finally, a more difficult (but potentially more useful) step would be to infer new tags, e.g., to infer *pot holes* from the note *Surface of road is very bumpy and full of pot holes*.
2. **Tag preferences to ratings.** Certain tags like *bike lane* and *rough* are used frequently and consistently in route preferences. This usage suggests a pattern for inferring ratings to suggest to a user: *You penalize blocks tagged gravel; there are 39 such blocks that you have*

not yet rated; would you like to rate them all "poor"?. Inferring ratings is particularly useful since the more users enter ratings, the better Cyclopath does at finding routes.

3. **Geoinference.** The geographic nature of Cyclopath raises unique and attractive opportunities for inferring tags: infer tags that have been applied to geographically nearby objects, and infer tags that have been applied to topographically connected objects (e.g., other blocks of the same trail or road).

All these methods are fallible. Thus, we also will experiment with mixed-initiative dialog models for users to monitor and approve/reject inferred tags. At least two possible models are attractive. *Basic wiki model*: Inferred tags are applied automatically. Users monitor for changes using normal wiki techniques (watch regions, recent changes) and then revert inappropriate tags. *Work Hints Model* [9]: Inferred tags are maintained as potential tag applications and presented to users through a visual interface. Inferred tags are applied only if a user approves.

8.2 Summary

We introduced tags into the Cyclopath geographic wiki for bicyclists. We chose the relatively rare design option of making tags owned by the community (broad scope) rather than individual users (narrow scope). We made this choice because it is consistent with the wiki model and to promote the goal of evolving a tag vocabulary that would be useful to the entire community. We also modified the Cyclopath route finder to let users use tags to fine tune their routing preferences. We studied the first year of tagging behavior in Cyclopath, finding that: (a) two quite distinct vocabularies emerged, one around geographic points and one around blocks in the transportation network; (b) while tags have evolved a distinct usage pattern, their use has some overlap with the use of other annotation techniques, and this raises interesting design opportunities; (c) user annotation behavior is highly specialized.

9. ACKNOWLEDGEMENTS

We thank the members of GroupLens Research, the Cyclopath team, and the Cyclopath user community. This work was supported in part by the NSF grants IIS 05-34692, IIS 08-08692, and IIS 09-64695 and by a GAANN fellowship.

10. REFERENCES

[1] S. L. Bryant et al. Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. In *Proc. GROUP*, 2005.

[2] E. Chi and T. Mytkowicz. Understanding the efficiency of social tagging systems using information theory. In *Proc. Hypertext*, pages 81–88, 2008.

[3] N. Garg and I. Weber. Personalized, interactive tag recommendation for Flickr. In *Proc. RecSys*, pages 67–74, 2008.

[4] S. Golder and B. Huberman. The structure of collaborative tagging systems. In *Proc. Information Science*, 2006.

[5] P. Heymann et al. Can social bookmarking improve web search? In *Proc. WSDM*, pages 195–206, 2008.

[6] S. F. King and P. Brown. Fix my street or else: Using the Internet to voice local public service concerns. In *Proc. ICEGOV*, pages 72–80, 2007.

[7] D. Millen et al. Dogear: Social bookmarking in the enterprise. In *Proc. CHI*, pages 111–120, 2006.

[8] K. Panciera et al. Lurking? Cyclopaths? A quantitative lifecycle analysis of user behavior in a geowiki. In *Proc. CHI*, pages 1917–1926, 2010.

[9] R. Priedhorsky et al. Eliciting and focusing geographic volunteer work. In *Proc. CSCW*, pages 61–70, 2010.

[10] R. Priedhorsky and L. Terveen. The computational geowiki: What, why, and how. In *Proc. CSCW*, 2008.

[11] A. Rivadeneira et al. Getting our head in the clouds: Toward evaluation studies of tagclouds. In *Proc. CHI*, pages 995–998, 2007.

[12] S. Sen et al. tagging, communities, vocabulary, evolution. In *Proc. CSCW*, 2006.

[13] S. Sen et al. The quest for quality tags. In *Proc. GROUP*, pages 361–370, 2007.

[14] S. Sen et al. Tagommenders: Connecting users to items through tags. In *Proc. WWW*, 2009.

[15] J. Thom-Santelli et al. Social tagging roles: Publishers, evangelists, leaders. In *Proc. CHI*, pages 1041–1044, 2008.

[16] K. Tso-Sutter et al. Tag-aware recommender systems by fusion of collaborative filtering algorithms. In *Proc. SAC*, pages 1995–1999, 2008.

[17] T. Turner et al. Picturing Usenet: Mapping computer-mediated collective action. *Computer-Mediated Communication*, 10(4), 2005.

[18] J. Vig et al. Tagsplanations: Explaining recommendations using tags. In *Proc. IUI*, 2009.

[19] H. Welser et al. Visualizing the signatures of social roles in online discussion groups. *Social Structure*, 8(2), 2007.

[20] H. Welser et al. Finding Social Roles in Wikipedia. In *annual meeting of the American Sociological Association, AllAcademic*, 2008.

[21] Y. Zhen et al. TagiCoFi: Tag informed collaborative filtering. In *Proc. RecSys '09*, pages 69–76, 2009.