Generating Personalized Spatial Analogies for Distances and Areas

Yea-Seul Kim
iSchool | DUB
University of Washington
yeaseul1@uw.edu

Jessica Hullman
iSchool | DUB
University of Washington
jhullman@uw.edu

Maneesh Agrawala
Computer Science Department
Stanford University
maneesh@cs.stanford.edu

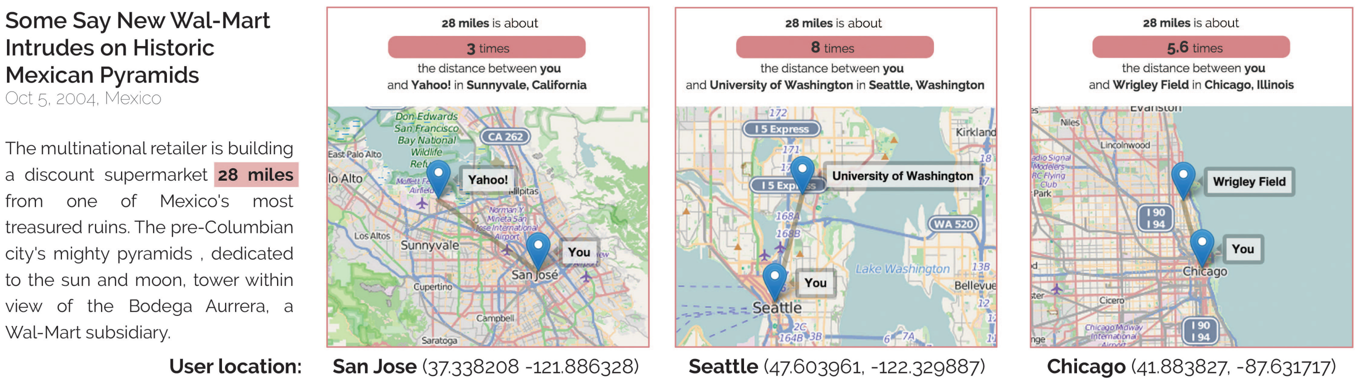

Figure 1. A text article containing a distance, alongside three personalized spatial analogies generated by our system for a user in San Jose, Seattle, and Chicago.

ABSTRACT
Distances and areas frequently appear in text articles. However, people struggle to understand these measurements when they cannot relate them to measurements of locations that they are personally familiar with. We contribute tools for generating *personalized spatial analogies*: re-expressions that contextualize spatial measurements in terms of locations with similar measurements that are more familiar to the user. Our automated approach takes a user's location and generates a personalized spatial analogy for a target distance or area using landmarks. We present an interactive application that tags distances, areas, and locations in a text article and presents personalized spatial analogies using interactive maps. We find that users who view a personalized spatial analogy map generated by our system rate the helpfulness of the information for understanding a distance or area 1.9 points higher (on a 7 pt scale) than when they see the article with no spatial analogy and 0.7 points higher than when they see generic spatial analogy.

Author Keywords
Distance; area; landmark; locator map; spatial analogy; information visualization.

CHI'16, May 07 - 12, 2016, San Jose, CA, USA

ACM ISBN 978-1-4503-3362-7/16/05 ... $15.00.
DOI: 10.1145/2858036.2858440

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION
People are frequently confronted with spatial measurements in news articles and other text documents. For example, a news article might mention the distance between the site of a plane crash and a nearby village (4.3 mi), or the square footage of a new prison (27,000 ft^2). Unfortunately, spatial references like these are difficult for people to understand and reason about. People often struggle with accurately interpreting numbers [37, 42] and find it challenging to judge distances and areas that they cannot easily relate to their experiential understanding of space [17, 21].

One strategy that authors use to directly relate new distances or areas to people's existing knowledge is to manually contextualize the measurement in terms of well known locations. For example, 4.3 miles might be conveyed as about the distance between the Empire State Building and the Brooklyn Bridge, or 420 acres as about half the size of Central Park. Such analogies can improve people's understanding of measurements by conveying the new information through a reference measurement that is well-known by the user [34, 35].

However, for a spatial analogy to aid cognition, readers must have prior knowledge about the locations and/or areas of the specific landmarks used in the analogy. While some landmarks may be recognized across populations, experiential knowledge of landmarks varies based on individualized factors like peoples' proximity to them [14, 21]. For example, the analogy of 4.3 miles as about the distance between the Empire State Building and the Brooklyn Bridge, or 420 acres as half the size of Central Park, only helps people that have good spatial knowledge of New York City. For a Detroit native a better analogy for 4.3 mi is about the distance from the Packard Plant to the Greektown casino. But, for an author to personalize such analogies to individual users would require knowing which landmarks the individual is familiar with – knowledge that is difficult to acquire.

We contribute tools for generating *personalized spatial analogies*: re-expressions that contextualize spatial measurements in terms of one or more locations with similar measurements but which are more familiar to a user. We first identify criteria for effective personalized spatial analogies, including the familiarity of the landmarks and how close the landmark's distance or area is to the target distance or area that is being contextualized. We present a set of automated tools that take a user's location and generate personalized spatial analogies for a target distance or area. Our approach generates the analogies by identifying landmarks that are similar to the target measurement but more familiar to user. We present an interactive personalized spatial analogy reader tool that tags distances, areas, and locations in a text article, then applies our automated algorithm to generate a personalized spatial analogy for each measurement. Our application presents the personalized spatial analogies for each target measurement via interactive maps (Fig. 1, 9).

We demonstrate the usefulness of our tools through a user study in which users are presented with text articles containing distances and areas. We find that users who view a personalized spatial analogy map generated by our system rate the helpfulness of the article content 1.9 points higher (on a 7 pt scale) for understanding the distance or area than when they see no spatial analogy and 0.7 points higher than when they see the article with generic spatial analogy.

RELATED WORK

Our work builds on prior research in three main areas; (1) studies in psychology and geography of how people reason about spatial information and the importance of landmarks in mental maps, (2) automated techniques for providing geographical context for locations, and (3) techniques for extracting landmarks from text and images.

How People Reason About Spatial Measurements

Cognitive geographers have described how mental models of space are metrically distorted: people rapidly become less spatially accurate and remember fewer details as locations get further away from the areas where they spend the most time (e.g., their home neighborhood) [14, 21, 45]. As a result, the locations that are encoded in the mental maps of different people may differ considerably. Thus, our approach for generating personalized spatial analogies relies on the distance of a user to a location as one proxy for their familiarity with that location.

"Landmarks" are salient reference nodes in a person's mental map of an environment that are distinctive from other points and therefore remembered more easily [33, 19, 14, 36, 43, 10]. Cognitive geographers describe several specific ways in which landmarks can be significant to people [33, 19, 14, 36, 10, 38, 29]. A landmark may be *generally* significant across a wide group of people for historical, spatial, or visual reasons [33, 10]. For example, the St. Louis Gateway Arch is significant for visual reasons as the largest arch in the world and taller structure than the surrounding St. Louis landscape and for historical reasons as a memorial to westward expansion. A landmark may also be *personally* significant to an individual based on his or her particular experiences with it [10]. For example, a neighborhood Starbuck's coffee shop might be significant to a resident of the neighborhood because she visits it regularly. Our approach considers both forms of landmark significance in selecting the landmarks to use in a personalized spatial analogy.

Though not aimed exclusively at providing geographical context, prior work has studied the benefits of manually created measurement analogies in which familiar classes of objects (e.g., a shopping mall, an adult male, etc.) are used to provide context for an unfamiliar measurement (e.g., 1 million ft^2, 80 kg) [34, 35]. Chevalier et al. provide a design study of concrete scales representations–visual re-expressions of complex measures–including strategies like reunitization, which we use [13]. Their work provides high level design guidance that influences our approach, including the importance of personal familiarity and an interpretable multiplier. However, our goal is to automate the process of generating measurement analogies that apply landmarks to re-express distances and areas using a database of landmarks that we extract from location-based services. To do so, we identify proxies that can be used to predict the personal familiarity of landmarks.

Providing Context for Spatial Measurements

Informed by how people think about space and landmarks, researchers have developed systems to generate maps that contextualize spatial measurements. Traffigram [28] uses isochronal cartography to generate maps that distort paths based on anticipated travel time. NewsViews [20] generates annotated thematic maps to accompany news articles that are relevant to spatially distributed data, such as a choropleth map of unemployment rates for an article about unemployment. Atlasify [27] generates maps that relate arbitrary text queries (e.g., "nuclear power") to spatial reference systems (e.g., a choropleth map depicting associations between "nuclear power" and countries in the world). Similar to these systems, we develop an automated approach to generate maps that relate spatial information to familiar concepts. However our approach addresses a different problem than the prior systems: how to produce spatial analogies of measurements in term of landmarks that are personally familiar to the user.

One simple strategy to help people understand locations is to present a "locator map": a map showing the locations that are mentioned in the text (Fig. 2). Studies have

Figure 2. A locator map.

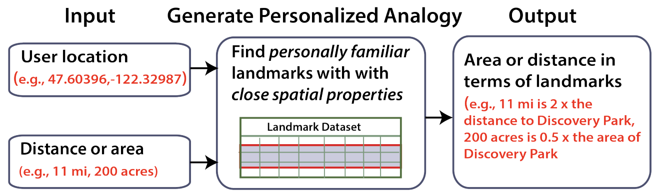

Figure 3. Diagram of our approach. The user's location, e.g., the center of Seattle, and a target distance or area are used to generate an analogy in terms of a landmark from a large landmark dataset. The goals in selection are to find a landmark that is familiar to the user and close to the target measurement in distance or area.

shown that adding locator maps to a text article improves understanding of the article by allowing for simultaneous textual and visual learning [24]. However, while locator maps are intended to help users better understand where countries and other locations are located, the goal of our work is to design an approach that can make spatial measurements, such as distances and areas easier for the user to understand.

Extracting Landmarks From Images and Text

Researchers have also developed techniques for extracting landmarks from large image and architectural data sets based on their distinguishing visual features, for example by using computer vision and structural information [8, 16, 23, 30, 39]. Other approaches mine landmarks from web text using text analysis techniques to determine landmark significance [43, 44]. We similarly leverage online sources to extract landmarks and assess their general familiarity across users. However, our interest is in applying the extracted data to generate personalized spatial analogies, rather than in developing a comprehensive set of landmarks.

APPROACH

Our goal is to generate *personalized spatial analogies*: expressions that contextualize a target spatial measurement (a distance or area) by relating it to spatial properties of one or more locations that are familiar to a user. Our approach takes two inputs: the user's geographical location and a target distance or area (Fig 3). These inputs are used to identify a landmark from a large landmark dataset that achieves two criteria: the landmark is personally familiar to the user and the landmark has a distance or area close to the target.

When the target measurement is a distance, our system generates a personalized analogy relating this distance to the distance between the user's location and a landmark. For example, an analogy might re-express a target distance of 11 miles as 2 times the distance from the user to a park that is located 5.5 miles from the user. When the target measurement is an area, the personalized analogy relates the target area to the area of a landmark. For instance, an analogy might re-express 200 acres as about half times the area of a 400 acre park near the user.

Criteria for Effective Spatial Analogies

We consider two criteria for effective spatial analogies that emerge from psychology research on landmark familiarity and numeracy.

Familiarity of the Analogous Spatial Property

For the spatial analogy to make the target measurement easy to understand, the spatial property of the landmark introduced in the analogy (i.e. the distance from user to landmark or the area of the landmark) must be familiar to the user. We use the term *personal familiarity* to refer to the user's level of familiarity with the analogous spatial property of one or more landmarks. While we cannot easily extract the personal familiarity of landmarks to a given user, personal familiarity is directly impacted by proximity [14, 21]. We therefore use proximity, or distance from the landmark's point location to the user's point location, as a proxy for personal familiarity. However, proximity alone does not fully capture a user's personal familiarity with a landmark. Some landmarks are recognizable to many people even if they are far away due to their cultural or historical significance, for example. We refer this type of popularity to as *general familiarity*. Our approach combines information on the user's proximity to the landmark with how generally familiar a landmark is across users to better capture personal familiarity.

Closeness of the Analogous Spatial Property to the Target

Our analogies relate a target distance or area to a multiple of the analogous spatial property of a landmark: e.g, 200 acres is 0.5 times the area of a 400 acre park near the user. We refer to the multiplicative factor for converting from the target measurement to the new measurement of the analogy (e.g. 0.5) as the *multiplier*. Research in number sense suggests that people are able to accurately reason about numbers between 1 and 3, though numbers between 3 and 10 also result in reasonably accurate inferences [7, 15, 18]. As numbers grow greater than 10, or dip below 1, the precision and ease with which people can reason about them decreases [7, 15, 18]. Hence, effective spatial analogies should avoid multipliers less than 1 and greater than 10.

OBTAINING LANDMARKS AND SPATIAL PROPERTIES

Our approach relies on a set of landmarks to generate personalized spatial analogies (Fig.3). We extract landmarks and their spatial properties from online sources as well as a proxy for general familiarity.

Identifying Landmarks and Their Spatial Properties

Yelp[1] is a crowd-sourced website containing reviews of thousands of businesses (e.g. restaurants, dry cleaners, etc.) and commonly visited places (e.g. monuments, parks, libraries). To seed our landmark database, we extract locations of business and places from Yelp (we limit to U.S. locations in our prototype implementation). However, since some service businesses are less likely to be associated with a well-known location (e.g., the office location of a tree removal service may not be known even by clients of the service), we do not include businesses in the Yelp service business categories (Event Planning & Services, Financial Services, Home Services, Local Services, Professional Services) in our database. For each of the non-service locations we store a name and location (in latitude/longitude coordinates) as generated by the Yelp API [4] in our seed database. With this procedure we obtain 242,194 Yelp locations as our seed set of landmarks.

[1]http://www.yelp.com/

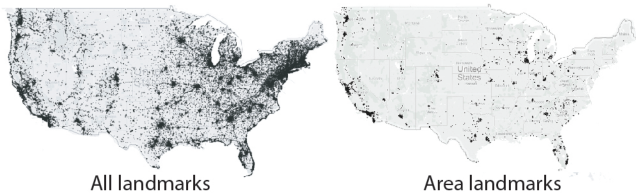

Figure 4. U.S. maps depicting all landmarks (left) and area landmarks (right) in our dataset.

To re-express and visually present area analogies we must also collect each landmark's area and footprint polygon. We first filter the landmarks to a subset of area landmarks—locations whose area is likely to be recognizable to people [26, 40]. While a person may be familiar with many locations near her (e.g., a gas station, an office building, a department store), it is not necessarily the case that he or she will be able to easily recognize the area of these places. For example, a person may find it difficult to imagine the area of a department store where he or she shops because the store has multiple floors or includes areas which they cannot access. On the other hand, some areas have clearer boundaries. We use a heuristic for determining area landmarks such as parks based on prior work that differentiates area landmarks from other landmarks [26, 40]. Specifically, in examining the categories of landmarks represented on Yelp we found that only a handful of categories describe landmarks whose area is likely to be recognizable: Parks, Stadiums & Arenas, Amusement Parks, Botanical Gardens, or Campgrounds. For area analogies we limit our dataset to the locations that fall in these categories, yielding 15,950 area landmarks. We also add U.S. states to this set of area landmarks, since the boundaries of US stages are likely to be recognizable to many users in the U.S.

For area landmarks, the polygon footprints of the landmarks are required to visualize them in the map. Yelp does not provide the polygon footprint for locations. Therefore, we extract this information from OpenStreetMap (OSM), a collaborative, open source mapping resource containing spatial features (nodes, ways, polygons) for locations around the world [25]. However, Yelp and OSM often use slightly different names to refer to the same location (e.g."SkyCity at the Needle", vs. "Sky City Restaurant"). We use a bottom-up approach to ensure that we accurately associate OSM areas and polygons with Yelp locations.

We first extract all continental U.S. locations from OSM for which the area property is not null (293,311 locations) and a footprint polygon is available. We then match the OSM locations to Yelp location by comparing the names of both locations using ngram similarity (a measure of the similarity between two strings [2] and the geographical coordinates of both locations (using the centroid coordinates for OSM locations). If the names of the two landmarks are close enough (i.e., above 0.5) and the coordinates of the two landmarks are close enough (i.e., the absolute difference is less than the average distance between matched landmarks: Latitude = 0.000169338147836, Longitude = 0.000172519196556), we consider the OSM landmark to be the same as the Yelp landmark and retain the area and polygon. This leaves 5,355 area landmarks. On a test set of 300 manually evaluated landmarks we obtain precision of 96.7% and recall of 68.4% using this matching process. Fig. 4 depicts the spatial coverage of the full set of landmarks (left) and the area landmarks (right; U.S. state area landmarks not pictured).

Assessing General Familiarity

We assess the general familiarity of each landmark by relying on how frequently it appears in photos in a large user-generated database. While other sources of familiarity information are available for Yelp landmarks (such as how many users have reviewed the landmark on Yelp), certain types of landmarks (e.g., restaurants, nightclubs) consistently receive more reviews than other common landmarks (e.g., monuments, bridges) [32]. We found that the number of photos taken of a landmark across a large set of users was less biased toward certain types of locations.

Specifically, for each of the landmark in our set of 242,194, we use the Flickr API [2] to find the total number of Flickr photos from Jan. 2010 to Jan. 2015 for which the landmark name appears in the photo tag and the photo was taken within 1 mile of the landmark's geographic coordinates. According to this measure of general familiarity, the most widely popular landmarks are Walt Disney World, Central Park, Golden Gate Park, Epcot Center, and the Washington Monument. Approximately 70% of the landmarks 169,128 do not appear in any Flickr photos, such as many small retailers, government offices, and motels. We do not omit these landmarks from the database, however, as there may be cases where they are very near to the user's location (i.e., high personal familiarity) and also close in area or distance to the target measurement. The remaining set of 73,117 landmarks has a mean photo count of 572.1 and median of 10. For area landmarks that are U.S. states, we record the photo count as the sum of the photo counts we obtain for all landmarks of the 242,194 in that state (mean: 408,552.5, median:227,952).

Our landmark dataset is available in a public repository [3].

GENERATING ANALOGIES USING ENERGY FUNCTIONS

Our approach takes a distance or area and the geographical coordinates of the user's location as input and uses an energy function minimization approach. The function considers a linear combination of terms that together capture the criteria for effective spatial analogies: closeness of the analogous property to the target measurement (*multiplier*), *personal familiarity* (proximity), and *general familiarity*.

Specifically, we define the energy function $E(\ell)$ of a landmark ℓ given user u and target distance t as:

$$E(\ell) = W_{pf}E_{pf}(u,\ell) + W_{gf}E_{gf}(\ell) + W_{mult}E_{mult}(t,u,\ell) \quad (1)$$

where $E_{pf}(u,l)$ is a function of the personal familiarity of l to u, $E_{gf}(\ell)$ is a function of the general familiarity of ℓ, and $E_{mult}(t,u,\ell)$ is a function of the multiplier required for the distance between u and ℓ to be equivalent to the t.

[2] https://www.flickr.com/services/api/

[3] https://bitbucket.org/yeaseulK/chi16-geospatial-analogy

We define the personal familiarity of a landmark to given a user $E_{pf}(u, \ell)$ as:

$$E_{pf}(u, \ell) = distance_v(u, \ell)$$

where $distance_v(u, \ell)$ is the Vincenty distance between ℓ and u. Vincenty distance is a slightly more accurate measure of distance than Euclidean distance as it accounts for the shape of the earth [46]. Our use of proxmity is based on the observation that landmarks that are closer to a user tend to be more personally familiar [14, 21]. For area landmarks including states, we use the centroid of the landmark's polygon to calculate proximity of a user.

We define the general familiarity of a landmark $E_{gf}(\ell)$ as:

$$E_{gf}(\ell) = \frac{1}{c_{Flickr}(\ell) + 1} \tag{2}$$

where $c_{flickr}(\ell)$ is the Flickr photo count for landmark ℓ. Since high photo count landmarks are likely to be more generally familiar and our goal is to minimize the energy, we take the inverse of the photo count, adding one to the denominator to avoid division by 0.

$E_{mult}(t, u, \ell)$ considers the multiplicative factor $mult(t, u, \ell) = \frac{t}{p}$ required to convert the analogous spatial property p of ℓ (the distance of u to ℓ for distance analogies or the area of ℓ for area analogies) to the target measurement t. We define $E_{mult}(t, u, \ell)$ as:

$$E_{mult}(t, u, \ell) = \begin{cases} 1/mult(t, u, \ell) & \text{if } 0 \leq mult(t, u, \ell) < 1 \\ 0 & \text{if } 1 \leq mult(t, u, \ell) < 3 \\ 0.1 & \text{if } 3 \leq mult(t, u, \ell) < 10 \\ \frac{mult(t,u,\ell)-10}{2} + 1 & \text{if } 10 \leq mult(t, u, \ell) \end{cases} \tag{3}$$

E_{mult} penalizes a landmark more heavily once the multiplier falls below 1 or if it exceeds 10, as the numbers 1 through 10 tend to be easier for people to reason with than numbers outside of this range [7, 15, 18]. We define the particular functions such that a multiplier just below 1 and just above 10 have the same penalty. When the multiplier of ℓ is between 1 and 3 (the most familiar numbers based on psychology studies of the reaction time required to recognizing different quantities [7, 15, 18]), we set E_{mult} to 0, and we assign a slight penalty when the multiplier is between 3 and 10 [7, 15, 18] (Fig. 5).

We use a weighted global criterion method [47] for multi-criteria optimization to find the values for the weights. Specifically, we repeatedly varied the inputs and weights while maintaining an a priori relative order of weights that we motivate based on prior geographical and numerical cognition research. Specifically, we weight personal as proximity the most, as a landmark that the user is not personally familiar with is unlikely to provide a useful distance or area reference. We weight the multiplier only slightly less than personal familiarity as proximity, as extremely large or small multipliers are also likely to make it much harder for a user to understand the analogy. We weight general familiarity the least since it does not necessarily predict whether the user is familiar with the spatial properties of the landmark.

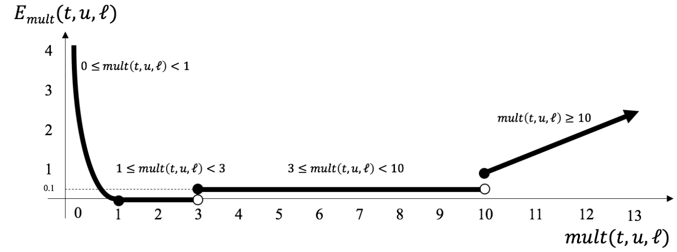

Figure 5. Plot showing penalty assigned by $E_{mult}(t, u, \ell)$ function by given magnitude of the multiplier.

Finally, we assign the following specific weights to each term:

$$E(\ell) = 5E_{pf}(u, \ell) + E_{gf}(\ell) + 4E_{mult}(u, \ell) \tag{4}$$

Depending on whether the target measurement is a distance or an area we apply the energy function to the set of all 242,194 Yelp landmarks or only the 5,355 area landmarks. We then select the lowest energy landmark for use in the spatial analogy.

Precision and Rounding

Retaining a high degree of precision (i.e., many decimal places) in reporting numeric values can make numbers more complex for users to interpret [31]. Rounding is a strategy that people use naturally to reduce complexity [11, 41]. While rounding can entail a loss of precision, hedges like "about" or "approximately" are often used to communicate that a value has been rounded [5].

To make multipliers easier to interpret while maintaining an upper bound on the loss of precision, we round multipliers to the nearest integer but adaptively retain extra decimal places as necessary until the rounded version is less than 5% different from the target measurement.

Results

Figures 6 and 7 show the top three personalized spatial analogies produced by our approach for various user point locations and a set of distance and area targets respectively. Many of the results include landmarks that are close to the user while still maintaining reasonable multipliers. As the input gets larger, multipliers grow larger than 10 to allow for landmarks that are more personally familiar to the user (i.e., higher proximity).

To assess the impact of the energy terms in our function independently, we generated similar tables for a set of distance and area inputs and user point locations but where one term in the energy function is turned off at a time. These results are available in our public repository [4].

APPLICATION

We develop an interactive personalized spatial analogy application that takes a geographical location inputted by the user and a distance or area and generates a personalized spatial analogy.

[4]https://bitbucket.org/yeaseulK/chi16-geospatial-analogy

	University of Washington (47.655328, -122.303493)			University of Michigan (42.278004, -83.738165)			Georgia Institute of Technology (33.775653, -84.396231)			University of California, Berkeley (37.871884, -122.258886)		
	Multiplier	Dist. (km)	Landmark	Multiplier	Dist.	Landmark	Multiplier	Dist.	Landmark	Multiplier	Dist.	Landmark
1 km	1.4	0.7	University of Washington	3.9	0.3	University of Michigan	2.2	0.5	Bobby Dodd Stadium at Grant Field	4.2	0.2	University of California Berkeley
	2.1	0.5	Husky Stadium	2.0	0.5	U of Mich School of Business	5.2	0.2	Georgia Tech Baseball	10.0	0.1	The Campanile
	3.7	0.3	Washington Huskies Football	1.0	1.0	Nichols Arboretum	1.2	0.8	The Varsity	1.2	0.8	California Theaters
5 km	6.9	0.7	University of Washington	9.8	0.5	U of Mich School of Business	3.7	1.4	Georgia Aquarium	52.1	0.1	The Campanile
	4.9	1.0	Neptune Theatre	4.9	1.0	Nichols Arboretum	3.5	1.4	Centennial Olympic Park	6.2	0.8	California Theaters
	5.3	0.9	University Village	7.6	0.7	AnnArbor.com	3.3	1.5	World of Coca Cola	4.8	1.0	University of California-Berkeley
10 km	9.7	1.0	Neptune Theatre	9.8	1.0	Nichols Arboretum	7.3	1.4	Georgia Aquarium	9.7	1.0	University of California-Berkeley
	4.9	2.1	Washington Park Arboretum	6.3	1.6	Michigan Stadium	6.9	1.4	Centennial Olympic Park	7.2	1.4	Lawrence Hall of Science
	4.0	2.5	Gas Works Park	9.1	1.1	The Blind Pig	6.6	1.5	World of Coca Cola	8.5	1.2	Berkeley High School
50 km	9.7	5.2	Space Needle	7.8	6.4	Matthaei Botanical Gardens	9.1	5.5	Candler Park	9.8	5.1	Berkeley Pier
	9.4	5.3	Seattle Center	6.7	7.5	Ann Arbor Cooks	9.5	5.3	Terminal West	9.7	5.2	Issues
	9.9	5.1	Seattle University	6.5	7.7	Putterz	8.3	6.0	The Earl	9.5	5.3	Cesar Chavez Park
100 km	9.4	10.7	Bellevue Botanical Garden	10.6	9.4	Eastern Michigan University	11.6	8.6	Lips	9.7	10.3	Lafayette Reservoir
	9.8	10.2	Alki Beach Park	8.8	11.3	Ypsilanti Automotive Heritage	8.8	11.3	Dick Lane Velodrome	9.5	10.5	Alameda Naval Air Museum
	9.1	10.9	Sound Transit	8.6	11.6	Saline District Library	10.7	9.4	Taste	9.7	10.4	St. George Spirits
500 km	23.2	21.5	King County	24.0	20.8	Canton Public Library	21.3	23.5	Stone Mountain Park	22.2	22.5	Golden Gate Park
	21.3	23.5	Seattle-Tacoma International Airport	26.1	19.2	Yankee Air Museum	20.6	24.2	Marietta Square	24.6	20.3	Golden Gate Bridge
	18.4	27.2	Flying Heritage Collection	20.5	24.4	Plymouth Ice Festival	22.1	22.7	Sweetwater Creek State Park	23.3	21.5	California Academy of Sciences
1,000 km	36.8	27.2	Flying Heritage Collection	28.5	35.1	Kensington Metropark	30.0	33.3	Kennesaw State University	34.0	29.4	Photos Marin
	33.1	30.2	ShoWare Center	30.5	32.8	Detroit Metropolitan Airport	30.8	32.4	Southeastern Railway Museum	32.9	30.4	San Francisco International Airport
	26.9	37.2	Snoqualmie Falls	31.0	32.3	Livonia Spree	35.6	28.1	Kennesaw Mountain	37.6	26.6	San Francisco Zoo

Figure 6. The top three personalized distance analogies generated by our approach for a range of targets and user locations.

	University of Washington (47.655328, -122.303493)				University of Michigan (42.278004, -83.738165)				Georgia Institute of Technology (33.775653, -84.396231)				University of California, Berkeley (37.871884, -122.258886)			
	Multiplier	Area	Dist. (km)	Landmark	Multiplier	Area	Dist.	Landmark	Multiplier	Area	Dist.	Landmark	Multiplier	Area	Dist.	Landmark
0.1 km2	2.7	0.04	1.94	Cowen Park	9.0	0.106	0.11	The Diag	0.9	0.11	1.44	Centennial Olympic Park	8.7	0.01	0.7	People's Park
	3.6	0.03	2.41	Louisa Boren Park	0.6	0.17	3.96	Eberwhite Woods	2.3	0.04	3.34	Orme Park	2.2	0.04	1.48	Codornices Park
	2.8	0.04	2.72	I-5 Colonnade Park	0.5	0.18	2.6	Miller Nature Area	6.2	0.02	3.53	John Howell Memorial Park	2.5	0.04	2.83	Lake Anza
0.5 km2	2.6	0.19	1.5	Ravenna Park	0.8	0.64	1.02	Nichols Arboretum	4.6	0.11	1.44	Centennial Olympic Park	11.1	0.04	1.48	Codornices Park
	3.5	0.14	1.78	Interlaken Park	0.8	0.67	3.56	County Farm Park	1.6	0.32	5.46	Candler Park	1.3	0.37	5.26	Cesar Chavez Park
	10.3	0.05	4.32	Cal Anderson Park	2.7	0.18	2.6	Miller Nature Area	11.6	0.04	3.34	Orme Park	1.7	0.30	6.19	Albany Bulb
1 km2	5.2	0.19	1.5	Ravenna Park	1.6	0.64	1.02	Nichols Arboretum	9.1	0.11	1.44	Centennial Olympic Park	2.7	0.37	5.26	Cesar Chavez Park
	7.0	0.14	1.78	Interlaken Park	1.5	0.67	3.56	County Farm Park	3.1	0.32	5.46	Candler Park	3.3	0.30	6.19	Albany Bulb
	0.7	1.38	3.4	Green Lake Park	5.4	0.18	2.6	Miller Nature Area	7.2	0.14	9.88	Medlock Park	1.1	0.93	7.1	Lakeside Park
5 km2	3.6	1.38	3.4	Green Lake Park	7.9	0.64	1.02	Nichols Arboretum	15.5	0.32	5.46	Candler Park	5.4	0.93	7.1	Lakeside Park
	7.1	0.70	7.94	Carkeek Park	7.5	0.67	3.56	County Farm Park	0.4	11.18	22.66	Sweetwater Creek State Park	2.7	1.85	9.42	Joaquin Miller Park
	3.9	1.27	9.58	Saint Edward State Park	6.2	0.80	6.56	Lillie Park	13.5	0.37	21.09	Graves Park	13.4	0.37	5.26	Cesar Chavez Park
10 km2	7.3	1.38	3.4	Green Lake Park	15.7	0.64	1.02	Nichols Arboretum	0.9	11.18	22.66	Sweetwater Creek State Park	5.4	1.85	9.42	Joaquin Miller Park
	7.9	1.27	9.58	Saint Edward State Park	15.0	0.67	3.56	County Farm Park	31.1	0.32	5.46	Candler Park	10.7	0.93	7.1	Lakeside Park
	5.3	1.90	8.86	Bridle Trails State Park	12.5	0.80	6.56	Lillie Park	3.1	3.18	29.84	Yellow River Park	3.3	3.01	15.32	Angel Island State Park
50 km2	4.0	12.42	19.63	Cougar Mountain Regional Park	10.7	4.69	25.29	Maybury State Park	4.5	11.18	22.66	Sweetwater Creek State Park	2.3	22.22	15.42	Briones Regional Park
	20.4	2.45	13.32	Marymoor Park	6.5	7.74	34.63	Willow Metro Park	15.7	3.18	29.84	Yellow River Park	16.6	3.01	15.32	Angel Island State Park
	26.3	1.90	8.86	Bridle Trails State Park	62.3	0.80	6.56	Lillie Park	12.9	3.87	47.79	Tribble Mill Park	12.3	4.05	22.48	Golden Gate Park
100 km2	8.1	12.42	19.63	Cougar Mountain Regional Park	21.3	4.69	25.29	Maybury State Park	8.9	11.18	22.66	Sweetwater Creek State Park	4.5	22.22	15.42	Briones Regional Park
	17.3	5.76	29.44	Lord Hill Regional Park	12.9	7.74	34.63	Willow Metro Park	31.4	3.18	29.84	Yellow River Park	24.7	4.05	22.48	Golden Gate Park
	40.9	2.45	13.32	Marymoor Park	36.6	2.73	67.04	Wildwood Preserve Park	13.8	7.23	52.54	Harbins Park	33.2	3.01	15.32	Angel Island State Park
500 km2	40.3	12.42	19.63	Cougar Mountain Regional Park	64.6	7.74	34.63	Willow Metro Park	44.7	11.18	22.66	Sweetwater Creek State Park	22.5	22.22	15.42	Briones Regional Park
	86.7	5.76	29.44	Lord Hill Regional Park	106.6	4.69	25.29	Maybury State Park	15.4	32.46	75.63	Hard Labor Creek State Park	45.6	10.96	67.44	Portola Redwoods State Park
	204.3	2.45	13.32	Marymoor Park	80.7	6.20	124.31	Pokagon State Park	69.1	7.23	52.54	Harbins Park	35.1	14.24	88.39	Calero County Park
1,000 km2	80.5	12.42	19.63	Cougar Mountain Regional Park	129.1	7.74	34.63	Willow Metro Park	30.8	32.46	75.63	Hard Labor Creek State Park	45.0	22.22	15.42	Briones Regional Park
	173.5	5.76	29.44	Lord Hill Regional Park	213.2	4.69	25.29	Maybury State Park	89.4	11.18	22.66	Sweetwater Creek State Park	70.2	14.24	88.39	Calero County Park
	353.7	2.83	42.41	Point Defiance Park	161.4	6.20	124.31	Pokagon State Park	50.5	19.81	112.66	Fort Mountain State Park	91.2	10.96	67.44	Portola Redwoods State Park

Figure 7. The top three personalized area analogies generated by our approach for a range of targets and user locations.

User Experience

Upon first using our application, a user is prompted to provide the address or spatial coordinates of a place that is highly familiar to them, such as their home address or a longtime previous residence. In browsing the web, when the user opens a text article, our application highlights distances and areas in the text, as well as all references to locations (including neighborhoods, cities, states, countries, and other locations, e.g., the Pentagon, University of Arizona, etc.). If the article references one or more locations, a locator map depicting the locations is presented to the right of the article (Fig. 8).

When the user clicks on any distance or area, a second map presenting a personalized spatial analogy for that measurement appears below the locator map (Fig. 9 lower left). When the user clicks on any single location in the text, the locator map is scaled to focus on the selected location, and a personalized area analogy map appears below the locator map (Fig. 9 right). The analogy map presents the selected location's area (e.g., a country) as a transparent overlay on the area landmark returned by our energy function approach.

If multiple locations (cities, neighborhoods, landmarks) are mentioned in the text, the user can select any pair of locations by clicking on them while pressing the shift key. The locator map is zoomed to show the two selected locations, and a personalized distance analogy map is shown below the locator map (Fig. 9, 8 center map).

System Implementation

Our application generates personalized spatial analogies in three steps (Fig. 10). First, a tagger identifies tags distances and areas that are referenced in the text article as well as the point location of the address provided by the user using the Google Maps Geocoding API [22]. When one of the tagged measurements is selected from an article the application generates the personalized spatial analogy for the target measurement using our energy function and landmark database. The top ranking spatial analogies returned by the energy function are then visualized in interactive maps.

Step 1: Tag Spatial Measurements and Locations

In step 1 we tag all distances and areas as well as locations that appear the article text. To identify distances and areas we match against a small set of regular expressions of the form:

[*number*][*optional separator*][*unit*]

where [*number*] matches any string of numeric digits with or without thousands separators, [*optional seperator*] matches

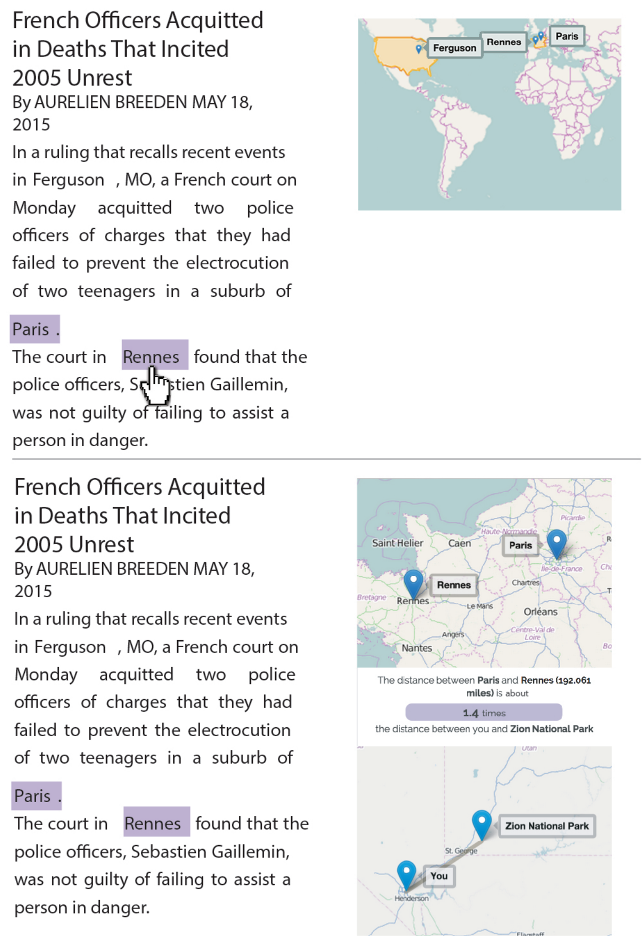

Figure 8. A user located in Las Vegas, NV (36.170864,-115.127571) interacts with our system as she reads a text article. By default, a locator map depicts locations referenced in the text. The user selects several cities in the article to see the distance between these locations contextualized through a personalized spatial analogy.

whitespace or dash (e.g., "-") characters, and [*unit*] matches any unit expression for distance or area (specifically, we match various spellings of *meters*, *kilometers*, *feet*, *yards*, and *miles* for distance and various spellings of the same units squared plus *acres* for area).

To identify locations in the text we apply named entity recognition (NER) tools on the article text. Specifically, we follow the approach of Gao et al. [20]. We first use Wikifier [12] to tag locations. Wikifier identifies named entities in text for which a Wikipedia article exists. For all tagged entities for which an article is found, we attempt to extract the geocoordinates and area from the article using the Wikipedia API [48]. We then use OpenCalais [3], a general-purpose NER tool, to remove any of the Wikifier locations that OpenCalais identifies as a *person* rather than a *location*. Those entities that remain are a good proxy for locations in the article. Gao et al [20] use a gold standard set of human-tagged locations in 47 articles to show that this location tagging approach achieves 92.7% precision and 42.6% recall.

This process results in a set of mentioned locations from the text along with their point locations and areas from Wikipedia. For all locations for which we successfully obtain an area, we attempt to obtain the geographical polygon for that location from OpenStreetMap.

The set of areas corresponding to identified locations and the set of distances between all pairs of identified locations become inputs for our energy functions.

Step 2: Generating Personalized Spatial Analogies

For each distance and area that results from step 1, we apply our energy function approach. The output of applying our energy function to a target distance or area includes the name, spatial coordinates, and multiplier for the top ranked landmark, as well as its footprint polygon for target measurements that are areas.

Step 3: Presenting Personalized Analogy

A renderer presents the results of step 2. The renderer highlights the tagged distances, areas, and locations the text. If the article contains one or more location references, a locator map depicting the locations is presented to the right of the text. When a user clicks on a highlighted distance, area, or location, they are presented with a map depicting the top ranked personalized spatial analogy for that measurement below the locator map (Fig. 8, 9).

Visualization: We implement all personalized distance analogy maps and the personalized area analogy maps for explicitly-referenced areas (e.g., 13 acres) using the Javascript-based library Leaflet [1]. Maps use the Web Mercator projection. The map extent is set using the default map extent approach provided in Leaflet, which centers the map on the centroid of all locations that are presented. All maps allow the user to zoom in and out using a +/- control in the top left of the map.

We construct the personalized area analogy maps for locations that are mentioned in the text using D3 [6]. D3 lets us use a different projection—the Lambert Azimuthal equal-area projection—that minimizes the distortion of areas between different locations on the globe. This projection ensures that the user can accurately compare the area of the selected location to that of the landmark returned as the top result by the energy function.

EVALUATION: USER STUDY

To evaluate our tools, we conduct a controlled user study in which participants are presented with text articles containing measurements. We hypothesize that users will find articles containing personalized spatial analogies generated by our system more helpful for understanding measurements than the most common baseline case in which an article does not contain analogies (H1). We also speculate that users will find articles containing personalized spatial analogies from our system more helpful than articles containing generic spatial analogies (H2). To generate generic spatial analogies, we applied our energy functions using a fixed location in New York City–the location of the Empire State Building–rather than the user's location. Generic spatial analogies are equivalent

Figure 9. An article containing spatial references for a user located in Seattle, WA (47.606831, -122.332427). When a user clicks on a distance or shift clicks on two locations (e.g., cities) in the text, a personalized distance analogy map is presented (lower left, center). When a user clicks on an area or single location, a personalized area analogy map is presented (top left, right).

to the analogies that our approach would generate for a user who is located at the Empire State Building. This form of analogy is similar to the type of general purpose analogy that a journalist at the New York Times might construct.

Study Design

We conducted our study using a between-subjects repeated measures design. We collected 10 text articles containing distances and areas and selected a single measurement in each article. For each article and measurement, we created one version of the article that includes a personalized spatial analogy map, and one "baseline" version of the article that includes no map. This comparison allowed us to assess H1, that users would find personalized spatial analogies more helpful for understanding measurements relative to a typical presentation of a text article. To assess H2 that users would find personalized spatial analogies more helpful for understanding measurements relative to a generic spatial analogy, we also create one version of each article that presents a "generic spatial analogy" which we generate by applying our energy function approach but with the fixed New York City location of the Empire State Building.

Study Procedure and Participants

For each article (trial), we highlight the selected distance or area. The participant is instructed to read the article. The participant then answers several questions. First, if the trial includes an analogy map, the participant is asked to:

- (*Landmark familiarity*): Rate how familiar you are with the landmark you see on the analogy map. (7 pt Likert)

This question serves as a manipulation check to confirm that the landmarks that appear in personalized spatial analogies are in fact more familiar to the user than those in generic analogies.

We also ask the participant to answer two additional questions (all trial):

- (*Helpfulness for understanding measurement*): How helpful is the content of the article (including text, images, and graphics) for helping you understand the size of the highlighted measurement? (7 pt Likert)
- (*Factors affecting helpfulness*): Briefly describe how the article content is or is not helpful for understanding the size of the highlighted measuremet. (1-2 sentence max).

Finally, for each trial we also require the participant to report the highlighted measurement using a text box. We use incorrect answers to this question to filter responses from participants who may not have paid attention.

We posted the study as a single HIT on Amazon's Mechanical Turk available to U.S. workers with an approval rating of 95% or higher. We ensured that the 10 trials assigned to each participant included at least one baseline version, one generic spatial analogy, and one personalized spatial analogy. Assignment of article versions was otherwise randomized and counter-balanced across participants. The HIT carried a reward of $3.00 and participants could earn an additional bonus of $1.00 if their responses to the attention verification questions are correct and their responses to the free text question were thoughtful.

Results

40 workers completed the task (400 trials). We removed responses for 1 trial in which the worker incorrectly identified the highlighted element. The mean time to complete the HIT was 26.3 minutes.

We first analyzed the results of our landmark familiarity manipulation check question to confirm that personalized spatial analogies did in fact use more familiar landmarks than generic spatial analogies. The mean perceived familiarity of the landmark for personalized spatial analogies was 1.8 points higher on a 7 point scale (μ=5.39, σ=2.53) than that for generic analogies (μ=3.62, σ=2.19; $t(232.92)$=-6.081, p <0.0001). (Baseline version users did not view a analogy map with a landmark so were not asked for this rating).

We next assessed H1: that users find personalized spatial analogies more helpful for understanding the measurements compared to the typical presentation of the baseline condition. We used an ANOVA ($F(2,347)$=30.43) followed by a TukeyHSD test to compare personalized spatial analogies to the baseline and to generic spatial analogies. The mean rated helpfulness of the content for understanding the size of the measurement was 1.9 points higher (μ=4.31,σ=2.14) for personalized spatial analogies than for the baseline (μ=2.41,σ=1.59; p_{adj} <0.001).

We then assessed H2: that users find spatial analogies more helpful for understanding the measurements compared to generic spatial analogies. The average helpfulness for the personalized spatial analogies condition was 0.69 points higher than the generic spatial analogies condition (μ=3.63,σ=1.98; p_{adj} <0.01). The TukeyHSD test also indicated that generic spatial analogies received higher ratings relative to the baseline (p_{adj} <0.001).

In describing factors that affected whether the content was or was not helpful, users of personalized spatial analogies frequently described personal experiences with the landmark they saw. For example, one participant mentioned that *'It helps to understand the distance because I have walked the trails around this landmark many times"*. On the other hand, multiple responses of users of generic spatial analogies expressed frustration toward the analogy when the landmarks were not familiar to the respondents: *"I'm from upstate New York and don't have much familiarity with the NYC area. Therefore it was difficult to contextualize the distance between the two places."* Other responses from users of generic analogies suggest that general familiarity can provide some help in allowing the user a rough guideline for judging the measurement size. For example, one user explained that *"Although I've never been to Central Park. I am a bit familiar with it from TV shows and movies. So the measurements given I can generally think of and what exactly they mean."*

DISCUSSION AND FUTURE WORK

Limitations

While self-reported helpfulness provides a useful signal of whether users find our tools helpful for understanding, further comprehension studies that elicit and test users understanding relative to their existing mental models is important for future work.

Future Work

The findings from our user study suggest that proximity of a user to a landmark effectively captures personal familiarity. This is a simple proxy, but our work showed how far this simple proxy could go in supporting the generation of useful personalized analogies. However, many other factors will

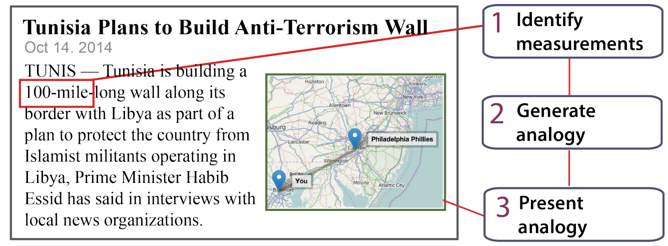

Figure 10. Our personalized spatial analogy reader application tags measurements and locations in the article. We use an energy function optimization approach to generate personalized spatial analogies, which are presented using interactive map visualizations. The example map depicts a personalized analogy for the distance 100 miles for a user at M.I.T. in Cambridge, MA (42.360051,-71.094214)

also contribute to personal familiarity, such as what types of locations a user spends time in or where people who are close connections to the user are located. Incorporating information from social media, navigation behavior will be another way to consider the other aspect of personal familiarity. Future work in automated personalized spatial analogies should explore additional factors beyond proximity and general familiarity that can be used to predict personal familiarity.

Other areas for future work include expressing analogies for distance using travel times in lieu of actual distances. For some people, travel time may be an easier proxy for understanding distance than a conventional distance measure [9, 28]. Our approach can be extended to present travel time information along with or instead of conventional distance measures.

While we believe that improving understanding of spatial measurements in a news reading context is a powerful application or our tools, our approach may also prove beneficial in other settings. For example, our personalized analogies may be useful in a digital learning setting for helping students develop a better understanding of geographical information. Other promising future applications include navigational applications, in which our tools might be used to deliver personalized distance or travel time information to help a user plan or navigate during a trip.

Familiarity and Multiplier Trade-off

Our study results show that the participants were more likely to find the analogies helpful when they accounted for proximity between the user and the landmark. When a landmark was not well known, as in generic spatial analogies, ratings were lower and participants often commented on their inability to relate to the size of the measurement. However, further study is needed to explore the trade-off between personal familiarity and the multiplier.

CONCLUSION

We contribute tools for generating personalized spatial analogies that utilize a user's location and a large landmark dataset to contextualize spatial measurements. We identify criteria for effective personalized spatial analogies and develop an energy function minimization approach for applying these criteria to select landmarks for which the analogous spatial property is close to the target measurement but which are more familiar to the user. We present an interactive application tool that tags distances, areas, and locations in a text article and presents personalized spatial analogies using interactive maps. We find that users who view a personalized spatial analogy map generated by our system rate the helpfulness of the article content higher than users who saw only the text article or a generic spatial analogy map.

ACKNOWLEDGEMENTS

This work was partially supported by a Tableau Fellowship.

REFERENCES

1. 2015. Leaflet JS. (2015). http://leafletjs.com/
2. 2015. NLTK Corpus Reader Documentation. (2015). http://www.nltk.org/_modules/nltk/corpus/reader/lin.html
3. 2015. Open Calais. (2015). http://new.opencalais.com/
4. 2015. Yelp API. (2015).
5. Susana Bautista, Raquel Hervs, Pablo Gervs, Richard Power, and Ra Williams. 2011. How to Make Numerical Information Accessible: Experimental Identification of Simplification Strategies. (2011).
6. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2301–2309. DOI: http://dx.doi.org/10.1109/TVCG.2011.185
7. Elizabeth M Brannon. 2006. The representation of numerical magnitude. *Current opinion in neurobiology* 16, 2 (2006), 222–229.
8. C Brenner and B Elias. 2003. Extracting landmarks for car navigation systems using existing GIS databases and laser scanning. *International archives of photogrammetry remote sensing and spatial information sciences* 34, 3/W8 (2003), 131–138.
9. Pat Burnett. 1978. Time cognition and urban travel behavior. *Geografiska Annaler. Series B. Human Geography* (1978), 107–115.
10. David Caduff and Sabine Timpf. 2008. On the assessment of landmark salience for human navigation. *Cognitive processing* 9, 4 (2008), 249–267.
11. Jamie I. D. Campbell. 2005. *Handbook of Mathematical Cognition*. Psychology Press.
12. X. Cheng and D. Roth. 2013. Relational Inference for Wikification. In *EMNLP*.
13. Fanny Chevalier, Romain Vuillemot, and Guia Gali. 2013. Using Concrete Scales: A Practical Framework for Effective Visual Depiction of Complex Measures. *IEEE TVCG* 19, 12 (2013), 2426–2435.
14. Helen Couclelis, Reginald G Golledge, Nathan Gale, and Waldo Tobler. 1987. Exploring the anchor-point hypothesis of spatial cognition. *Journal of Environmental Psychology* 7, 2 (1987), 99–122.
15. Stanislas Dehaene, Elizabeth Spelke, and Lisa Feigenson. 2004. Core Systems of Number. *Trends in Cognitive Sciences* 8, 7 (2004), 307–314.
16. Birgit Elias. 2003. Extracting landmarks with data mining methods. In *COSIT*. Springer, 375–389.
17. John Eliot. 2002. About spatial intelligence: I. *Perceptual and Motor Skills* 94, 2 (April 2002), 479–486. DOI: http://dx.doi.org/10.2466/pms.2002.94.2.479
18. Lisa Feigenson, Stanislas Dehaene, and Elizabeth Spelke. 2004. Core systems of number. *Trends in cognitive sciences* 8, 7 (2004), 307–314.
19. Nathan Gale, Reginald G Golledge, William C Halperin, and Helen Couclelis. 1990. EXPLORING SPATIAL FAMILIARITY. *The Professional Geographer* 42, 3 (1990), 299–313.
20. Tong Gao, Jessica R Hullman, Eytan Adar, Brent Hecht, and Nicholas Diakopoulos. 2014. NewsViews: an automated pipeline for creating custom geovisualizations for news. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 3005–3014.
21. Reginald G. Golledge and Robert J. Stimson. 1996. *Spatial Behavior: A Geographic Perspective*. The Guilford Press, New York.
22. Google. 2015. Google Maps Javascript API. (2015). https://developers.google.com/maps/documentation/javascript/examples/geocoding-simple
23. Floraine Grabler, Maneesh Agrawala, Robert W Sumner, and Mark Pauly. 2008. *Automatic generation of tourist maps*. Vol. 27. ACM.
24. Jeffrey L. Griffin and Robert L. Stevenson. 1994. The Effectiveness of Locator Maps in Increasing Reader Understanding of the Geography of Foreign News. *Journalism & Mass Communication Quarterly* 71, 4 (Dec. 1994), 937–946. DOI: http://dx.doi.org/10.1177/107769909407100417
25. Mordechai (Muki) Haklay and Patrick Weber. 2008. OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing* 7, 4 (Oct. 2008), 12–18. DOI: http://dx.doi.org/10.1109/MPRV.2008.80
26. Stefan Hansen, Kai-Florian Richter, and Alexander Klippel. 2006. Landmarks in OpenLSa data structure for cognitive ergonomic route directions. In *Geographic information science*. Springer, 128–144.
27. Brent Hecht, Samuel H Carton, Mahmood Quaderi, Johannes Schöning, Martin Raubal, Darren Gergle, and Doug Downey. 2012. Explanatory semantic relatedness and explicit spatialization for exploratory search. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 415–424.

28. Sungsoo Ray Hong, Yea-Seul Kim, Jong-Chul Yoon, and Cecilia R Aragon. 2014. Traffigram: distortion for clarification via isochronal cartography. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 907–916.
29. Alexander Klippel and Stephan Winter. 2005. Structural salience of landmarks for route directions. In *Spatial information theory*. Springer, 347–362.
30. Kengo Koiso, Takehisa Mori, Hiroaki Kawagishi, Katsumi Tanaka, and Takahiro Matsumoto. 2000. Infolod and landmark: Spatial presentation of attribute information and computing representative objects for spatial data. *International Journal of Cooperative Information Systems* 9, 01n02 (2000), 53–75.
31. Manfred Krifka. 2002. For verbosity and precision1. *Sounds and systems: studies in structure and change. A festschrift for Theo Vennemann* 141 (2002), 439.
32. Michael Luca and Georgios Zervas. 2013. Fake it till you make it: Reputation, competition, and Yelp review fraud. *Harvard Business School NOM Unit Working Paper* 14-006 (2013).
33. Kevin Lynch. 1960. *The image of the city*. Vol. 11. MIT press.
34. Alejandra Magana, Sean Brophy, and Lynn Bryan. 2012. An Integrated Knowledge Framework to Characterize and Scaffold Size and Scale Cognition (FS2C). *Int'l J. of Science Educ.* 34, 14 (2012), 2181–2203.
35. Alejandra J Magana. 2014. Learning strategies and multimedia techniques for scaffolding size and scale cognition. *Computers & Education* 72 (2014), 367–377.
36. Daniel R. Montello. 1997. The Perception and Cognition of Environmental Distance. In *Direct Sources of Information,Proc. Int. Conf. on Spatial Information Theory: A Theoretical Basis for GIS*. 297–311.
37. John Allen Paulos. 1988. *Innumeracy: Mathematical Illiteracy and Its Consequences*. Macmillan.
38. Martin Raubal and Stephan Winter. 2002a. *Enriching wayfinding instructions with local landmarks*. Springer.
39. Martin Raubal and Stephan Winter. 2002b. *Enriching wayfinding instructions with local landmarks*. Springer.
40. Kai-Florian Richter and Alexander Klippel. 2005. A model for context-specific route directions. In *Spatial cognition IV. Reasoning, action, interaction*. Springer, 58–78.
41. Bruce M. Ross and Trygg Engen. 1959. Effects of round number preferences in a guessing task. *Journal of Experimental Psychology* 58, 6 (1959), 462–468. DOI: http://dx.doi.org/10.1037/h0049112
42. Charles Seife. 2010. *Proofiness: The Dark Arts of Mathematical Deception* (1 ed.). Viking Adult.
43. Taro Tezuka and Katsumi Tanaka. 2005. Landmark extraction: A web mining approach. In *Spatial information theory*. Springer, 379–396.
44. Taro Tezuka, Yusuke Yokota, Mizuho Iwaihara, and Katsumi Tanaka. 2004. Extraction of cognitively-significant place names and regions from web-based physical proximity co-occurrences. In *Web Information Systems–WISE 2004*. Springer, 113–124.
45. Barbara Tversky. 2003. Structures of mental spaces how people think about space. *Environment and behavior* 35, 1 (2003), 66–80.
46. Thaddeus Vincenty. 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review* 23, 176 (1975), 88–93.
47. Chankong Vira and Yacov Y Haimes. 1983. *Multiobjective decision making: theory and methodology*. Number 8. North-Holland.
48. Wikipedia. 2004. API: Main page. (2004). https://www.mediawiki.org/wiki/API:Main_page