Research article

Mopsi location-based service

Pasi Fränti*

Machine Learning, School of Computing, University of Eastern Finland, FINLAND

*   Correspondence: Email: pasi.franti@uef.fi.

Academic Editor: Chih-Cheng Hung

Abstract: Mopsi is a location-based platform for storing photos and GPS tracks of the users. It allowed user to share their data on-line with real-time user location, communicate with other users, share the data in Facebook, browse the collected photos and tracks on map, perform searches and ask recommendations. Mopsi was operational from 2009 to 2020. This paper documents the history of Mopsi, its main functionalities, research achievements, and the collected data.

Keywords: location-based services; LBS; Mopsi; ubiquitous computing; mobile platform

1.   Introduction

Location is a small piece of information about the user or their data. This extra information has emerged many innovative applications like navigation, sport trackers, logistics optimization, ubiquitous recommendation, and location-aware social networks.

The seed for the development was the integration of GPS technology into mobile phones pioneered by Benefon, a Finnish mobile phone producer, who launched their first GPS-phone in 1999 followed later by Nokia, iPhone and other manufacturers. The expectations were high, and it took time for the technology to mature. Nowadays, all smart phones have mobile navigation by-default and very sophisticated *location-based services* (LBS) are emerging.

Mopsi[1] is a location-based platform created for prototyping research results developed in two research projects in the university of Eastern Finland during 2008–2014. Mopsi supported collecting

---

[1] http://cs.uef.fi/mopsi/

geo-tagged photos and trajectories collected by mobile phone. The data collection remained active until 2020. In total, Mopsi has 64k photos and 11k GPS trajectories collected by 203 users.

Browsing the existing data is still possible and many of the main functionalities are still operational. The data is heterogenous and useful for testing research ideas. Smaller subsets have been extracted from the full data to study specific research questions.

In this paper, we document the Mopsi platform, its history, the main functionalities implemented in the platform, and summarize the data collected. Mopsi research activities has continued with various intensity. We overview the most important research achievements so far. They represent a horizontal view of the kind of research problems appearing in location-based applications.

Figure 1 presents the original three main functionalities in Mopsi: location-based search, photos, and trajectory collections (called routes). Photo collections is the most used part in Mopsi. The collected trajectories raised many new research ideas. The other main parts are the search and recommendation system, web content mining that supports the search, social networking aspects, and location-based games (LBG). We will go through these themes one-by-one.
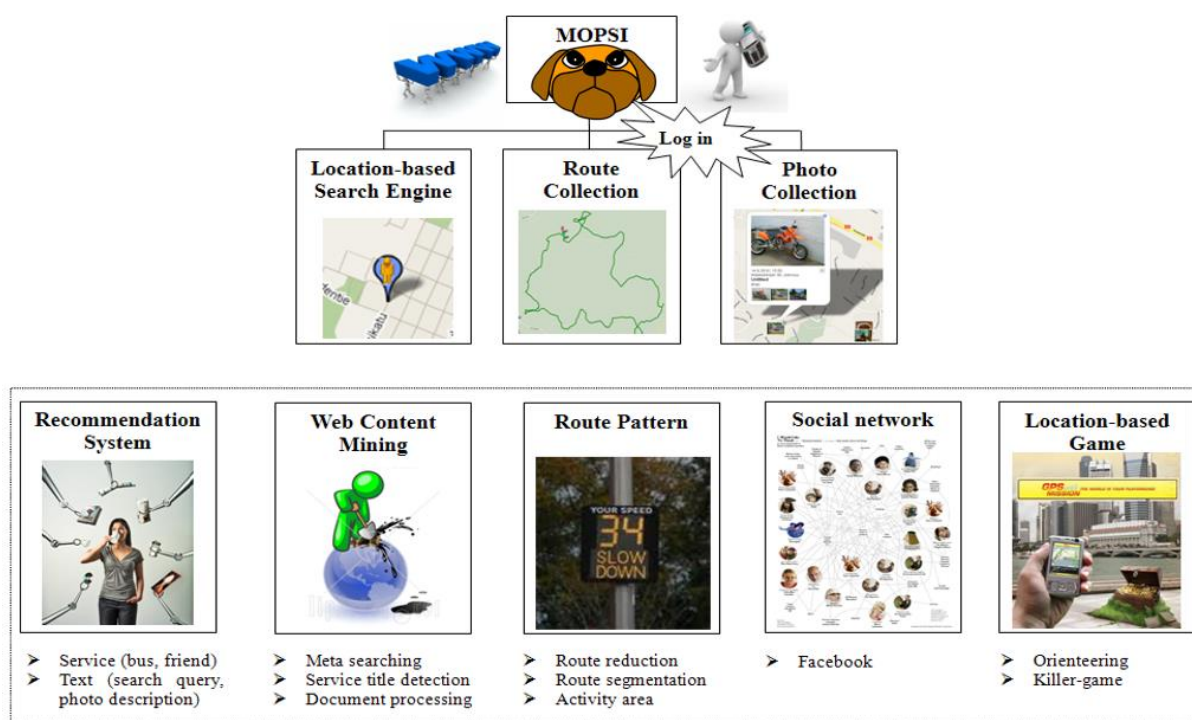

Figure 1. Overview of Mopsi. The top part represents the main services offered by Mopsi. The bottom part is the vision by Qinpei Zhao in 2011 about what themes the future research at that time was expected to cover. All ideas but the Killer Game were indeed studied and later implemented in Mopsi.

Figure 2 summarizes the data collected in Mopsi[2]. The size is not huge, but it is heterogenous, rich, collected in real world environment outside of laboratory condition. Most data are geo-tagged photos and GPS tracks. Other data consist of O-Mopsi games and the Mopsi services, which are

---

[2] http://cs.uef.fi/mopsi/data/

business and other services manually created (mostly in Joensuu area) with some level of quality control. Photos and tracks were controlled by removing known test data and retain only real users.
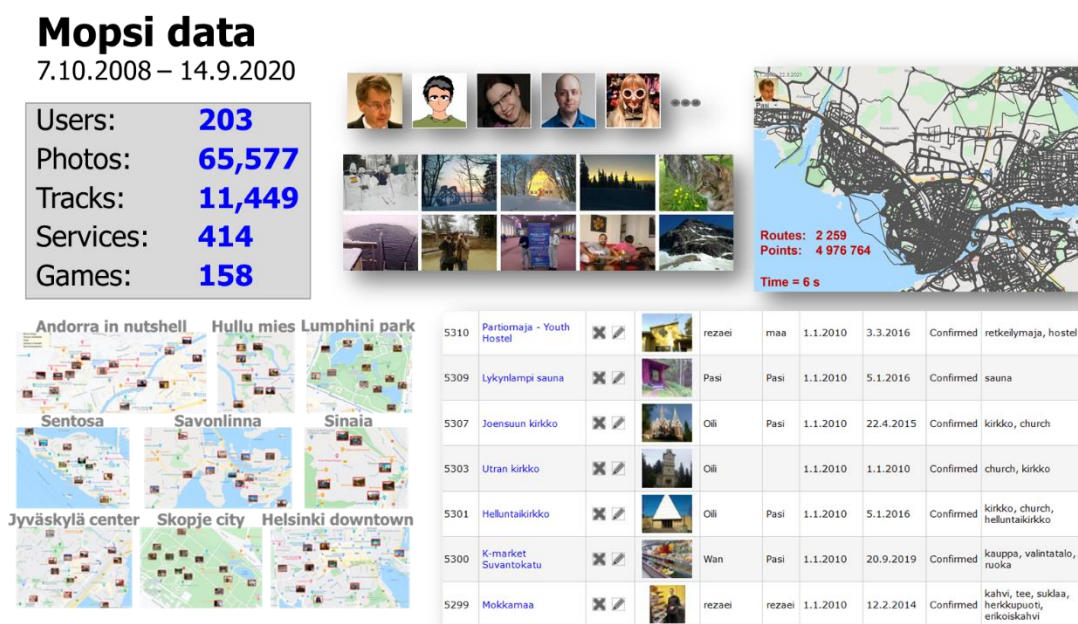

Figure 2. Summary of the collected in Mopsi between Oct-2008 and Sept-2020.

2. Location-based services

Data can be attached with a small piece of extra information: location. We define *location-based service* (LBS) as a service where this extra piece (location) plays an important role.

In one end, there are *cartography* and *geographic information systems* (GIS) where the location is everything. The data simply does not exist without location information. A point in two-dimensional space (x,y) is just a point, but when it represents a geographic location, it becomes meaningful. A sequence of coordinates can form movement trajectories, and polygons encapsulating areas. When attached with an attribute like *blue* or *water*, it will become meaningful in cartography.

Significant geographic location is called *point-of-interest* (POI). It can be a business outlet like cafeteria, major landmark, or any place with something recognizable object. The points can have also other data like name, purpose, opening hours, or a free-form text description what is there.

The location can also be a larger entity. Four level of hierarchy from the most precise (exact geo-coordinate) to the least precise (country) is shown in Figure 3 according to [1]. Business outlets usually give their location by postal address, which was the key source of location-aware information on webpages [2]. Postal codes, city (municipal), or country level aggregations are also used for statistical analysis when comparing different regions.
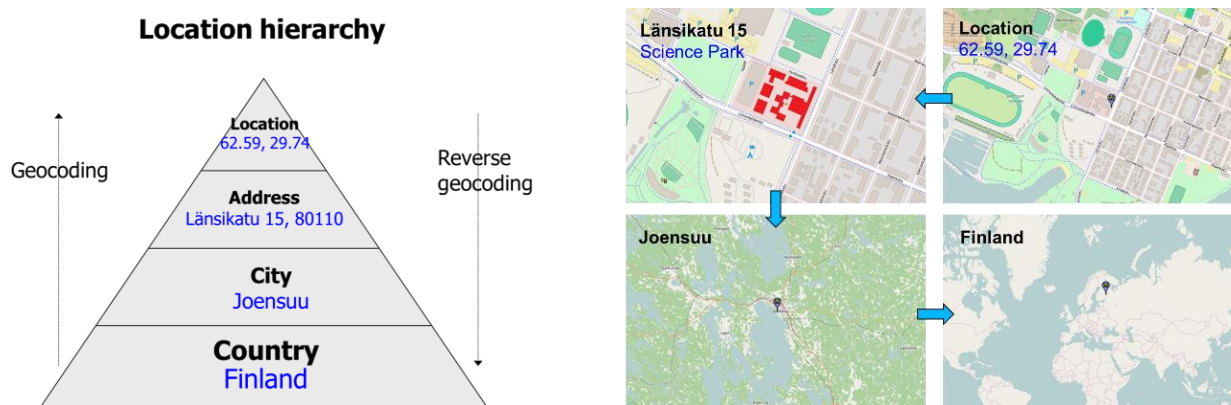
Figure 3. Representing location can vary from a very detailed level (single GPS point) to high level (country) [1].

2.1. Location data

Research have focused on visits (where we move), movement trajectories (how we move), and social-network data (who we meet) [3]. The visit data is often called *check-ins*, which can be manually input by the user or automatically concluded from the user movements. The location and the related semantic information are called *geo-social environment* in [4].

A typical LBS application is a recommender system. The recommended items can be popular places concluded from the check-ins and movement trajectories [5]. New less-known tourist destinations were also discovered in [6] based on content-similarity of photos. In addition to the location, the recommended items can be the photos themselves and user trajectories [7,8].

A research question is to determine which data is relevant to the user. According to [9], relevance depends on four factors: *content*, *time*, *location*, and *social network*. The content can be the information attached to the POI (what is there), free-form text description, tags given by the user, or automatically extracted semantic content of a photo. Time can refer to recency (more recent are more relevant), season (skiing photos are more relevant in winter), and opening hours. Location relevance depends mostly on the distance to the user or places they visit regularly.

2.2. Trajectory data

Trajectories can be formed of dense GPS points, sparse cell data connections, or simply a sequence of user check-ins. The frequency of the points varies from high frequency data like in the Joensuu subset (2s) of the Mopsi trajectories [10] and Chicago campus data (4s) [11] to sparse datasets like Berlin (42s) and Athens (61s) [12].

A classical research problem is route planning. User can be recommended by point-to-point trips optimized for criteria like pedestrian health [13], air pollution exposure [14], or selecting the best attractions for a sight-seeing tour. The recommendation result can also be an entire route travelled by someone else [7]. The tour can be given as a treasure hunt game where players need to visit given without any predefined route or visit order to follow [15].

Trajectory data have also been used for more challenging research problems like extracting entire road network from the GPS tracks [10,16], or jointly with data from aerial images [17]. The motivation

can be to create a new road network from scratch, update an existing one by adding new road segments [18], or creating personalized networks for a given user or user group.

2.3. Social network

Social networks in LBS can be explicit friendship networks, but also infer automatically from communication between users, or their joint appearance in the same place at the same time [19]. In location-based social networks, the most popular research topics according to a literature review [5] are location recommendation, route planning, friendship prediction based on user similarity, influence maximization, community detection, event detection, and privacy protection.

According to the *small-world effect*, we can reach everyone by six steps. Experiments with Twitter data gave even a smaller number, 3.43 [20]. We are theoretically well connected. However, having 261 Facebook-friends does not mean we have 261 real friends. If we want to share information, we can expect only a fraction of our connections to re-distribute it (say 1%). Friends of a friend will re-distribute even less likely (say 0.1%). Simple calculations in Figure 4 indicate that we could reach 1,305 users in this way.

The probability of sharing the information depends on the connection strength. The study in [21] indicated that semantic similarity creates a stronger connection than location-similarity. However, when the participants were asked to estimate their similarity to other users, the results reflected more about perceived similarity (or admiration) rather than real similarity. The same bias was confirmed with Twitter data [22]. In brief, the efficiency of spreading information depends on the strength of the connection, and how influential the user is.
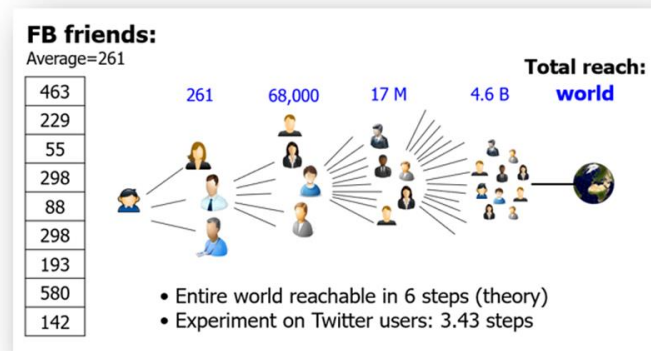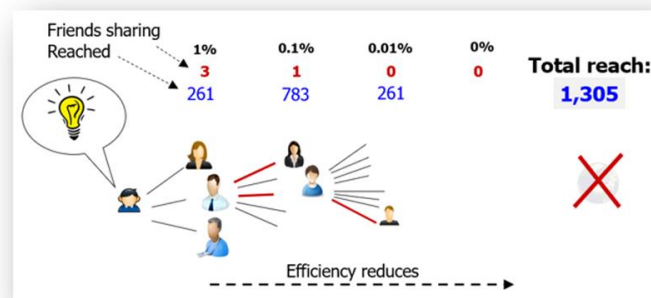

Figure 4. Small-world effect does not guarantee that we would reach everybody in world.

2.4. Privacy

Sharing location has security concerns. The most obvious is a robber sneaking to your home when you are travelling. In Mopsi, the privacy concern was not addressed, and the user-consent was simply made as an agreement to share all data by-default for the sake of research. Active users were mainly post-graduate students. With wider popularity, the privacy issue should have been considered. A simple solution is to limit the visibility of data.

There are three main solutions to address privacy concerns according to [23]: (1) decisions whether to share the data at all (decision could also be automated); (2) anonymization of the data; and (3) data obfuscation by adding spatial perturbation to the locations. Sharing only with friends reduces the threat but do not completely solve it. In our health care project, two alternative solutions were taken. In [24], the algorithm used the exact data in the optimization but showed only randomized locations in the visualization of the result. In [25], the locations were aggregated to the postal code level, and excluding entries with five or less data points.

Smart phones have recently started to support approximate location instead of the exact location. It makes the data obfuscation easier but reduce reliability of some applications. However, it was found in [26] that a significant share of the functionalities of applications were not working at all with the approximate location. The privacy issue is therefore not easy to solve.

3. Mopsi history

In 2000–2002, our research group [3] worked on two research projects (REALMAP and DYNAMAP) with companies like *Arbonaut*, *Benefon*, and *Kata-electronics*, focusing on vector and raster maps and their applications. The general assumption then was that vector maps would soon replace raster maps. Contrary to this expectation, the raster format remained the primary format for a long time on mobile devices. Our solution suggested the dynamic use of compressed raster images with buffering to support the offline use of maps [27]; see Figure 5.

Mopsi started as a side-track of this research, initially as an idea for a location-based search engine. We wanted to utilize two separate data sources: a location-based service database and a classical search engine. Two Indian interns [1] first tested and formalized the idea (see Figure 6) two years even before the launch of GoogleMaps in 2005.

After that, three subsequent master's students continued the work as a side-track with small development steps. Contrary to my expectations, the third student successfully implemented the search engine idea with promising results. The prototype reached a 50% recall rate by finding about half of the services in the Joensuu city area by simple keyword search. This success motivated the launch of two research projects.

The first project, *Mobile Location-based Applications and Internet* (MOPSI) was ongoing from 2008 to 2011. The second project, *Mobile Personalized Information Systems* (MOPIS), had a less innovative name to differentiate it from the first project. The name of the first project was designed so that its acronym could match the word *"Mopsi"*[4]. Since then, Mopsi has remained the name of the

---

[3] https://cs.uef.fi/ml/

[4] Mopsi, in Finnish, means pug (a dog). In Romanian, it is Mops.

platform and related mobile Apps, see Figure 7. Accordingly, the faces of a Mopsi dog were used; see Figure 8.

Due to the low activity in the early years, it took seven years from the publishing of the idea of location-based search [1] to when the prototype was published [28,29], and seven more years before the concept was fully documented in detail [2]. The search engine remained in Mopsi from the beginning to the end. Despite its importance, the projects focused on other ideas for practical reasons. The main reason was that we did not expect to compete with big companies like Google, with far more resources. The services of Google Maps and its competitors have developed far beyond what we ever implemented in practice.

The research focus changed more to user-collected data: trajectories and geotagged photos. Mopsi's main functionalities consisted of these two (called *Mopsi Photos* and *Mopsi Routes*) and the search engine called *Mopsi Search*. The latter remained the same as described in [29] but was later extended to include a recommender system [7,8].

Mopsi was implemented in 2008–2014 as a web tool where users can browse their photos using a timeline view and on a map. Other services in Mopsi included bus schedules for two cities, Joensuu and Kouvola, as well as many other hidden functionalities. Some had a separate App, while only the developers knew where to find others. In this paper, we document the most important ones.
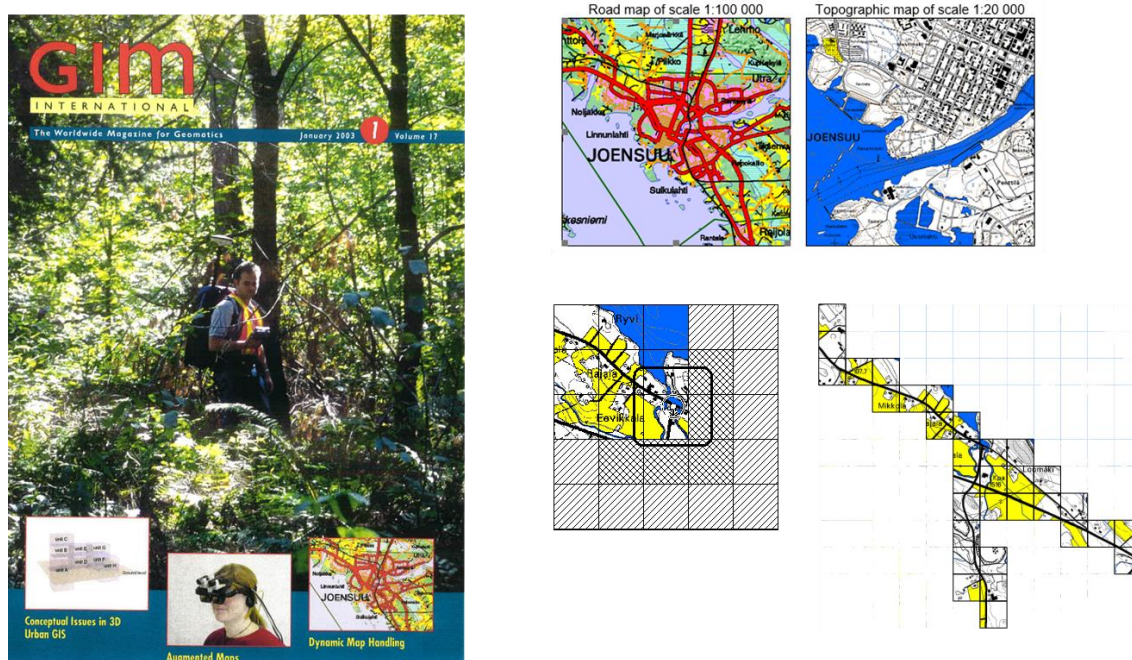

Figure 5. Early research resulted in a dynamic map image handling system suitable for mobile devices. The solution is documented as a summary paper in *GIM International* magazine [27].
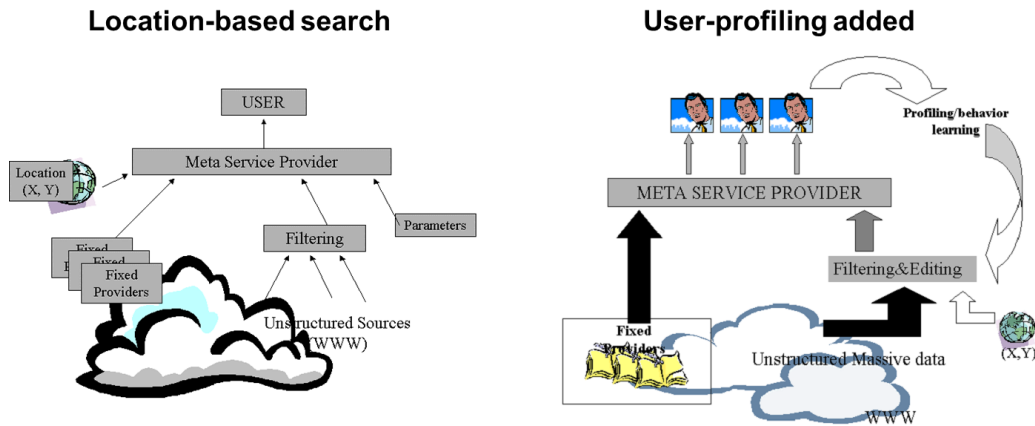
Location-based search

User-profiling added

Figure 6. The first draft of a location-based search engine in [1] consists of two parts: a geotagged database of services (fixed providers) and a crawling-based search engine (unstructured sources).
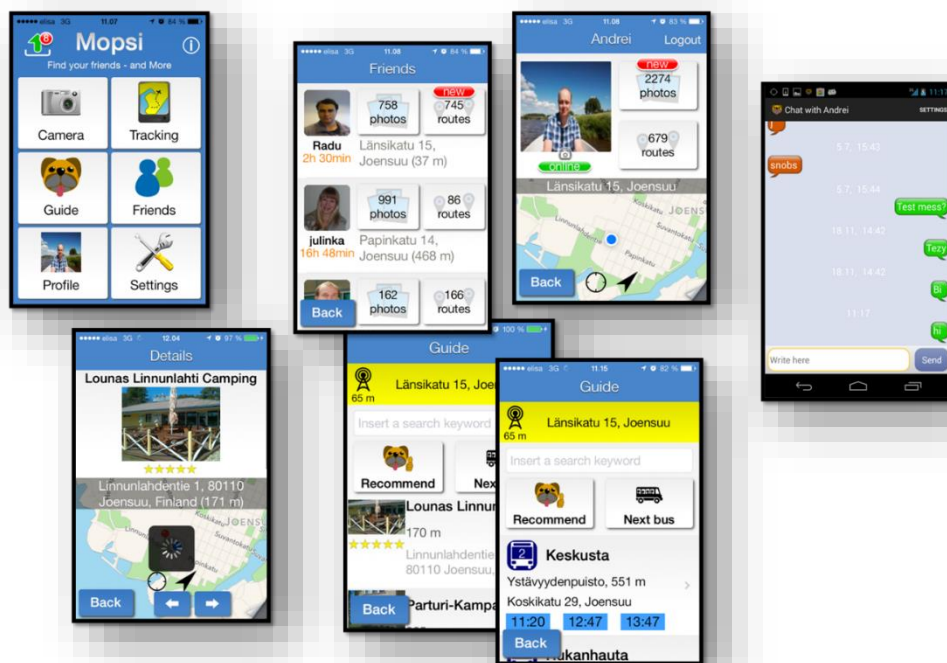

Figure 7. Mopsi was implemented in all major mobile operating systems at the time: Symbian (Nokia), WP (Microsoft), Android, and iPhone. The main screen stabilized to the six-button view, but other parts varied between the versions, depending on who implemented it.

Figure 8. Some Mopsi logos and faces in Mopsi. The original logo was created by Andrei Tabarcea and other versions by Radu Mariescu-Istodor.

4. Photo collection

There are many photo collection sites on the web that support geo-tagging. The specialty of Mopsi is that having a location is not an option—every photo must have it. Photos were uploaded by Mopsi mobile applications using positioning by default. If the location was missing, some versions allowed the user to add it manually (and tune later), but by default, no photo was accepted without some location. As a result, the database includes 64,660 photos, of which only less than 50 lack coordinates at position (0,0) on the map.

Mopsi web includes three main functionalities for viewing the photo collections: (1) browsing on timeline view (Figure 9), browsing on a map (Figure 10), and a typical single-image viewer capable of moving to the next and previous images. Both timeline and map views include a clustering component: a single image is presented by a thumbnail and clusters by thumbnail, with a number in a circle indicating how many photos are in the cluster.

Both views have two main functionalities. Clicking the photo opens the selected cluster by re-adjusting the time period (timeline view) or map scale (map view) to match the photos inside the selected cluster. Clicking the number circle starts classical photo viewer mode with the photos of the selected cluster. Real-time clustering adapts to panning and scale changes made by the user. When opening a cluster, the photos inside the cluster are re-clustered to optimize the selected view.

The cluster interface is an efficient way to search a given photo by its location. For example, we can find the photo of Singapore Flier by four clicks starting from the map view of Asia (Figure 11). Each click narrows the area: Southeast Asia, Malaysia Peninsular, and Singapore. Then, clicking Singapore Downtown reveals five photos with Singapore Flier at the top.

The clustering is done by a grid-based algorithm with an additional merge step. It has been tailored for efficiency and has public API, including server and client-side variants [30]. The clustering on the timeline view is done by a simple k-means algorithm.
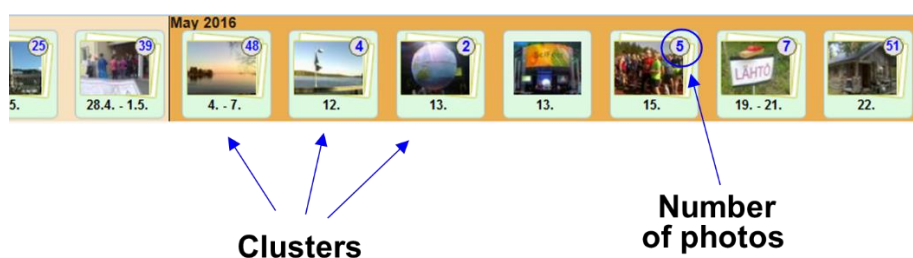

Figure 9. Timeline view to browse Mopsi photos. The number in the circle indicates how many images are inside each cluster. Clicking the number shows the content in a photo viewer. Clicking the thumbnail moves the cluster hierarchy to one level deeper.
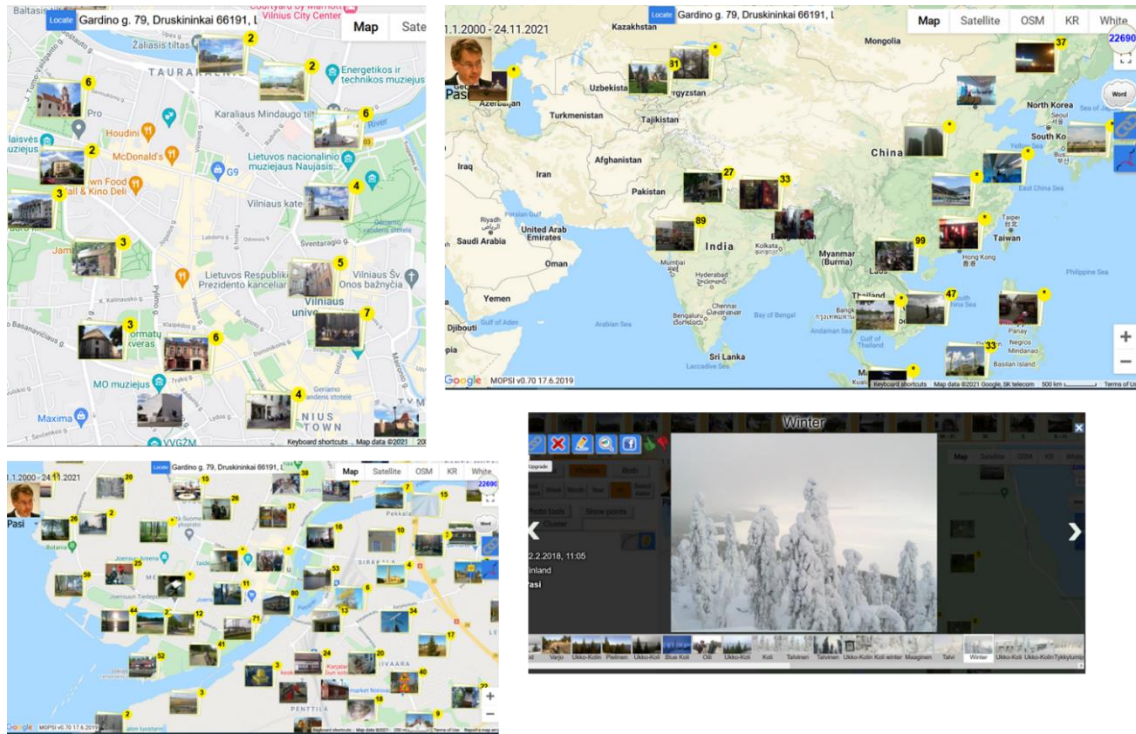
Figure 10. Examples of photo clusters on the map view and the photo viewer. Clicking a photo opens the selected cluster. Clicking the number starts the photo viewer.
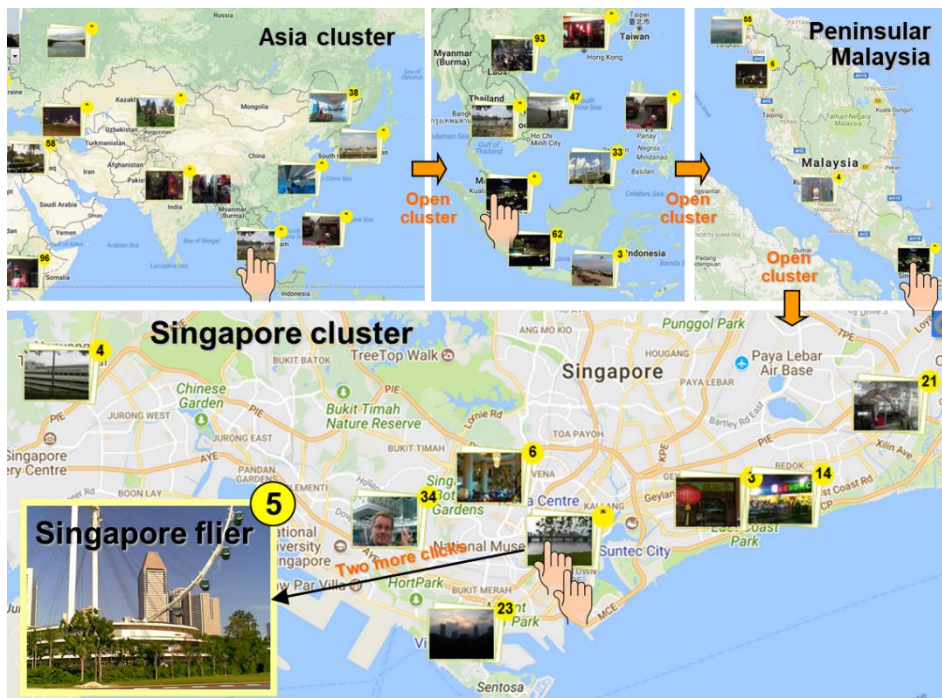

Figure 11. Clicking the cluster provides an efficient way to find a desired photo.

5. Mopsi routes

Mopsi Routes is the second main functionality in Mopsi. It allows users to track their movements and store them for later analysis. The possibility of sharing via Facebook existed for years. While there are many other sports tracker applications, such as Strava, Mopsi includes many sophisticated research-oriented functionalities that other sports trackers lack.

The route collection (2008–2019) includes 6,779 tracks collected by 51 users in the real world, see Table 1. User movements were mainly walking, running, riding a bicycle, and driving a car. The data is heterogeneous and contains GPS noise due to the lower quality of smartphones at that time. The data has served as a testbed for many research ideas. Figure 12 shows collected tracks.

Several subsets of the overall collection have been extracted for research purposes:

- **Mopsi Routes 2014**: a subset of 6,779 routes (7,850,387 points) recorded by 51 users between 19.7.2008 and 31.12.2014. [31]
- **Mopsi Routes 2019**: a subset of 2,484 routes (3,409,812 points) recorded by ten users between 31.3.2018 and 31.3.2019, mostly in Joensuu, Finland. (this paper)
- **Joensuu trajectories**: 45 tracks of a single user (Pasi) between 16.11.2014 and 25.4.2015 in Joensuu downtown. [10]
- **Mopsi segments**: There are 355 road segments in Joensuu downtown, including 1222 individual tracks extracted from the previous subset [32].

Table 1. Summary of the Mopsi routes data.

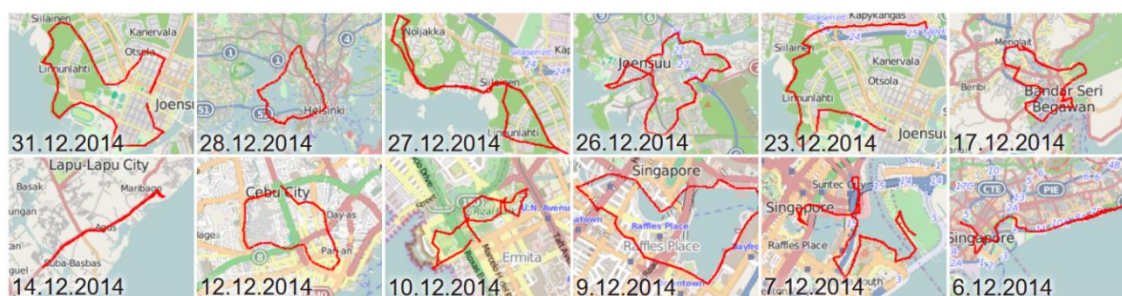| Data summary:	
Routes	6,779
Points	7,850,387
Kilometers	87,851
Hours	4,504
Collection details:	
Who	51 Mopsi users
When	19 July 2008–31 December 2014
How	Mobile phones with various movement types
Issues	Includes plenty of GPS errors

Figure 12. Examples of GPS tracks in Mopsi from 2014.

Mopsi implemented a multi-scale route management system [33] to allow fast access to the data on a map, see Figure 13. After a trajectory had been uploaded to the system, a reduced version was

pre-calculated by an efficient linear-time polygonal approximation algorithm [34]. Users can access this data by zooming and panning via the Google Maps interface. Mopsi can simultaneously show more than 2,000 tracks on a map in less than 10 seconds; see Figure 14. Most other platforms do not usually even allow the showing of multiple tracks.
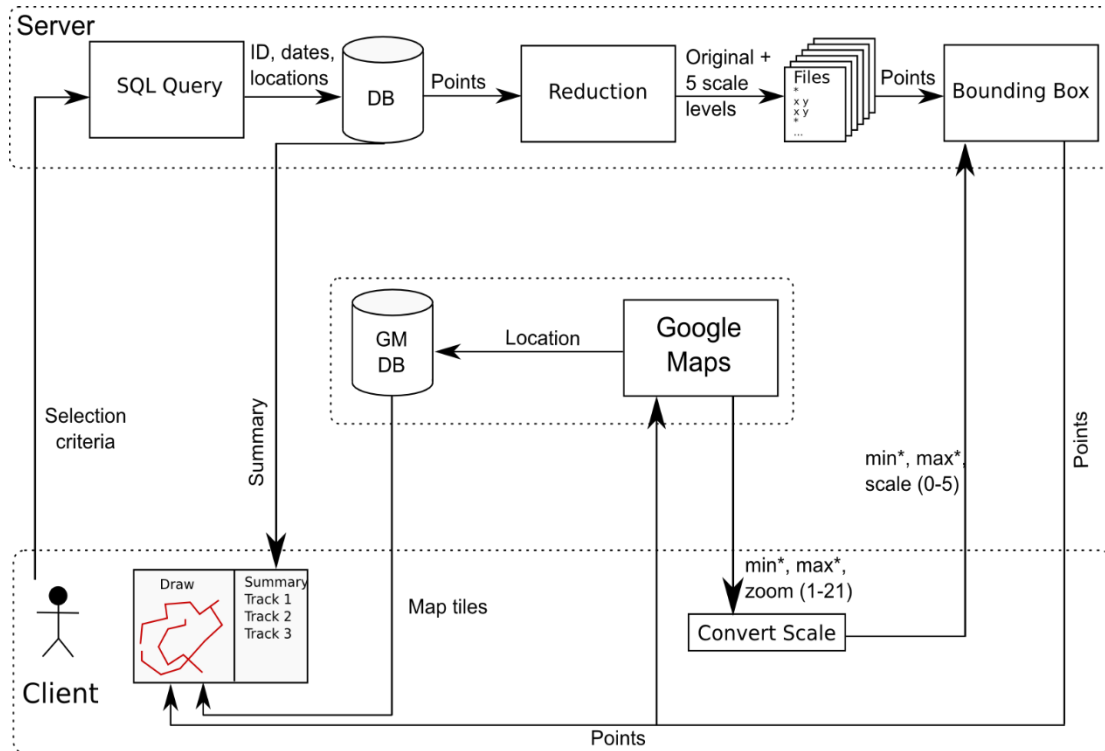

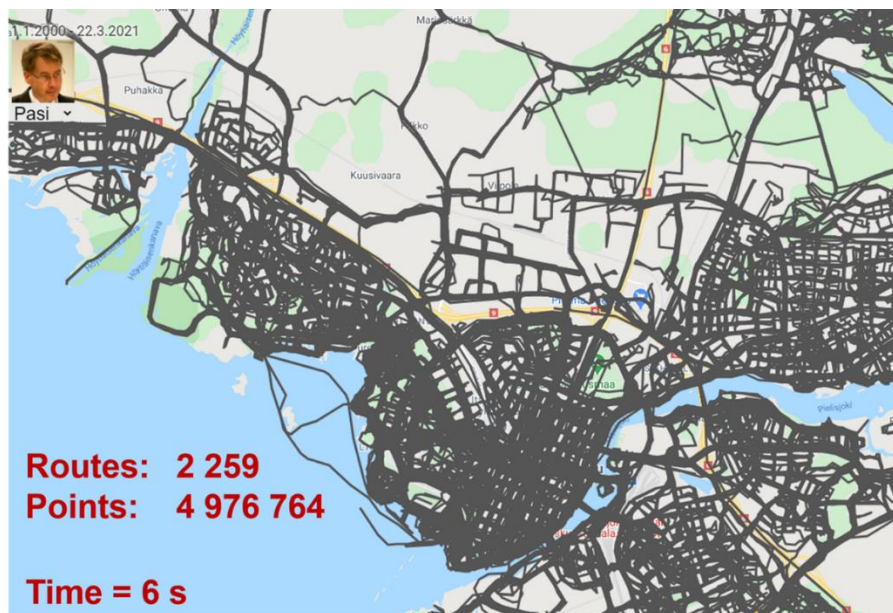**Figure 13.** Real-time processing of the tracks in Mopsi [33].

Figure 14. Real-time visualization of the collection of the most active user (Pasi).

The user selects a target date and the number of days before and after the target date. Mopsi then calculates a bounding box of all the retrieved tracks. It defines the map area that will be shown on screen. The corresponding map scale is selected, and reduced versions of the tracks of this scale are inputted into Google Maps. The tracks are then shown on the map. There is also a list of the tracks on the left side of the screen with their key information (date and length); see Figure 15. The interface has a few shortcut buttons for most typical queries: most recent, last week, last month, last year, and all.
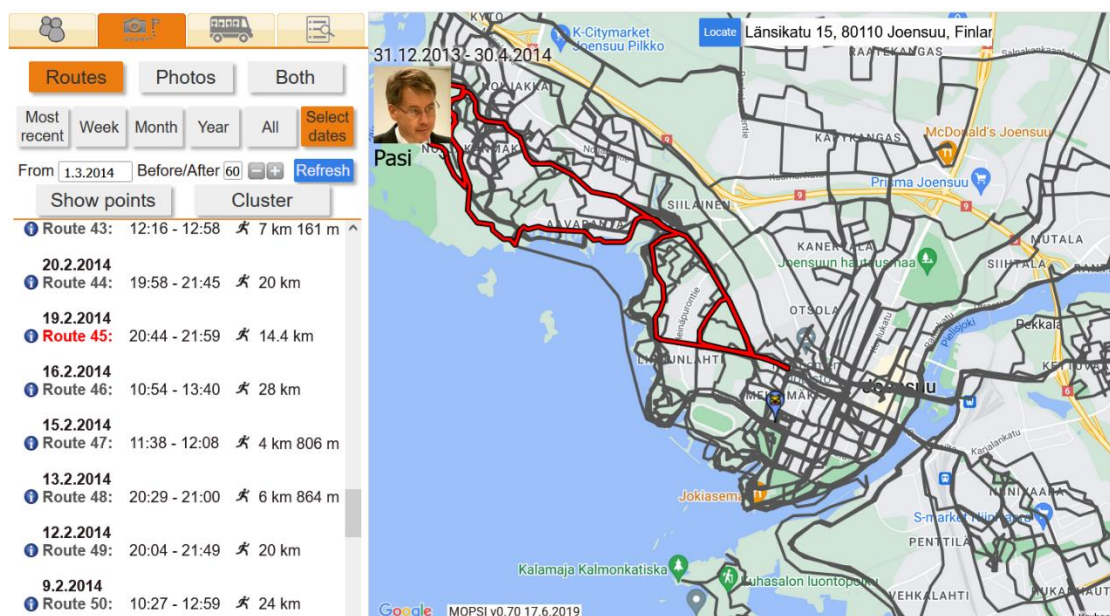

Figure 15. Search results for Pasi's route on 1 March 2014, +/- 60 days, are shown as a list (left) and on the map (right). The selected route is emphasized by red color. It can be opened in an analysis tool by pressing the blue info button.

Mopsi has two more advanced search functions: *similarity search* and *gesture search*. They use a grid-based similarity measure called C-SIM [31], which is fast and robust to noise and changes in sampling rate; see Figure 16. Two improved variants were also developed but not integrated in Mopsi: a hierarchical variant called HC-SIM [32] and a context-aware variant [35] for situations where routes are separated by obstacles such as rivers and buildings; see Figure 17.

The similarity search finds all tracks with a similarity value greater than 0. They can be browsed on a map. Statistical comparison of two tracks is also possible. The gesture search [36] can be activated by pressing the CTRL button and drawing a free-hand sample (by mouse). The system converts the drawing into a GPS track, which is the input to the same similarity search using C-SIM as the similarity function, see Figure 18.

Mopsi has many other analysis tools as well. Some were designed for a need; some were ideas of creative minds and used the Mopsi data for testing. The most useful ideas were integrated into the operational part of Mopsi. Others have public API on the web or separate web pages for demonstrative purposes. Some may have a hidden place in Mopsi where only the developers know how to use it. The following ones are worth mentioning:

- Move type detection [37]
- Novelty detection [38]
- Roundness detection [39]
- Road network extraction [10]
- Segment averaging [32,40,41]
- Fast travel distance estimation [42]
- User similarity (GPS biometrics) [43]
- User similarity (location history) [44]
- Destination prediction [45]

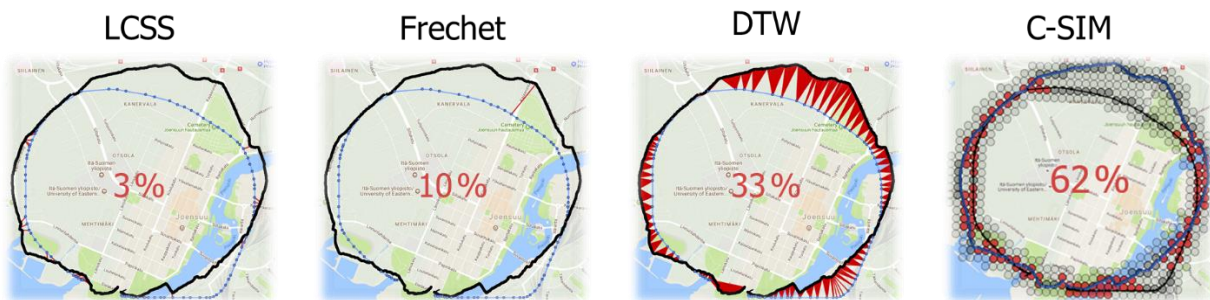These will be reviewed in more detail in a follow-up paper [46].

Figure 16. Common similarity measures for trajectories include longest common subsequence (LCSS), Fréchet, and dynamic time warp (DTW). Mopsi uses cell similarity (C-SIM) due to its low time complexity, and better tolerance to noise and changes in sampling rate and noise.

Measure	Correlation
HC-SIM	0.84
C-SIM	0.72
IRD	0.52
LCSS	0.45
EDR	0.37
Hausdorff	0.32
ERP	0.21
DTW	0.11
Euclidean	0.09
Discrete Frechet	0.05

$$HC\text{-}SIM\,(A, B) = \frac{1}{L}\sum_{i=1}^{L} C\text{-}SIM(A, B, 0.005 \times 2^{i-1})$$
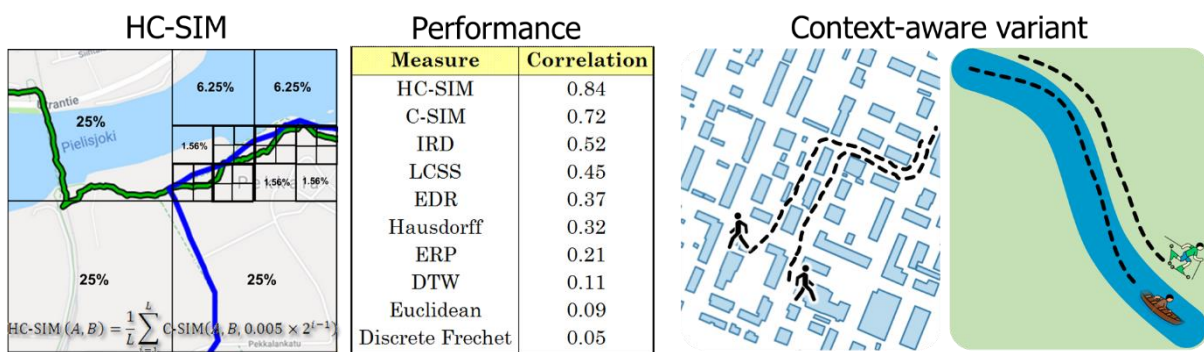
Figure 17. Improved variants of the grid-based similarity measure. The hierarchical variant sums up the C-SIM scores of each level by doubling the weight at every higher scale. It has the highest correlation to human evaluation [32]. The context-aware variant adds a penalty to trajectories separated by buildings or rivers [35].
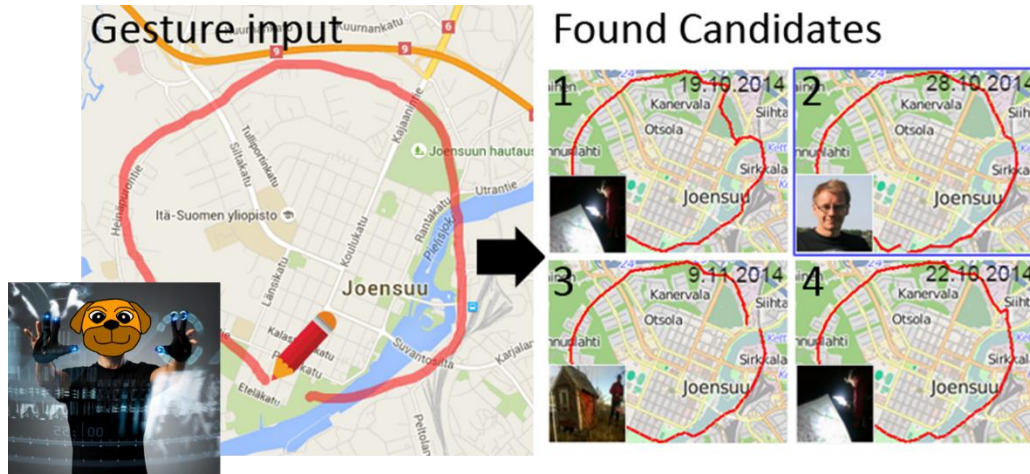
Figure 18. Gesture search [36] uses a free-hand drawn curve as input and searches using the grid-based similarity measure (C-SIM). An example of searching round trips around Joensuu downtown resulted in four candidates.

6. Other functionalities

Next, we outline selected other functionalities implemented in Mopsi. Some can be found in the Mopsi menus, others in less obvious places, and some might no longer work.

6.1. Mopsi search and recommendation

Mopsi search was the original research goal. It is a special case of a typical web search but limited to pages attached to a location. Geotagging was first considered, but we soon realized that street addresses are much more common on service web pages [2]. By default, search results are given only to places near the user location determined by mobile positioning (Mopsi Mobile) or by pointing the mouse on the map (Mopsi Web).

Once the focus of Mopsi changed from being a search engine to user-collected data, the search was also extended to include user-collected photos based on their text description. The third data source was the Mopsi service database of 414 entries. They are mainly in Joensuu downtown, which was covered reasonably well but is now outdated. The service data initiated research on web mining, including title extraction [47,48], keyword extraction [49,50], representative image extraction [51], website classification [52], and string similarity [53,54].

Mopsi Search was later extended to become a recommender system. In fact, the search button has the text *"Recommend"* by default. It changes to *Search* only when the user starts to enter some text into the keyword box, see Figure 19.

In the case of Mopsi services, the recommender system uses three factors to calculate relevance: search history, distance, and popularity of the service; see Figure 20. The keywords of the recommended services were analyzed by counting how many times they were used by others in total, recently, in nearby areas, and by the user himself. The weighted sum of these was used as the relevance factor. In the case of user-collected photos, the same function was used, but the service keywords were replaced by words used in the text description of the photos.

Mopsi web search is fully documented in [2], including extracting addresses and other relevant information from the web pages. The recommender system was introduced in [8] based on an early draft. Future improvements were considered using social networks [21] and recommending local events [55]. A prototype of the event recommendation was in Mopsi (http://cs.uef.fi/mopsi/events/), and a visual analysis tool is available here: https://cs.uef.fi/mopsi/events/analysis.html.
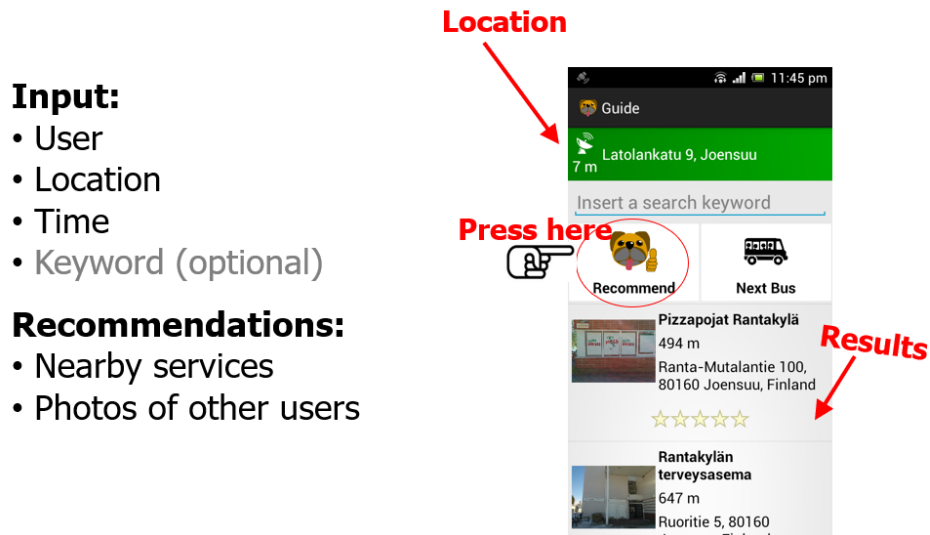
Figure 19. Mopsi search is actually a recommender until the user starts to input the search phrase. After that, the text with the button changes from *Recommend* to *Search*.
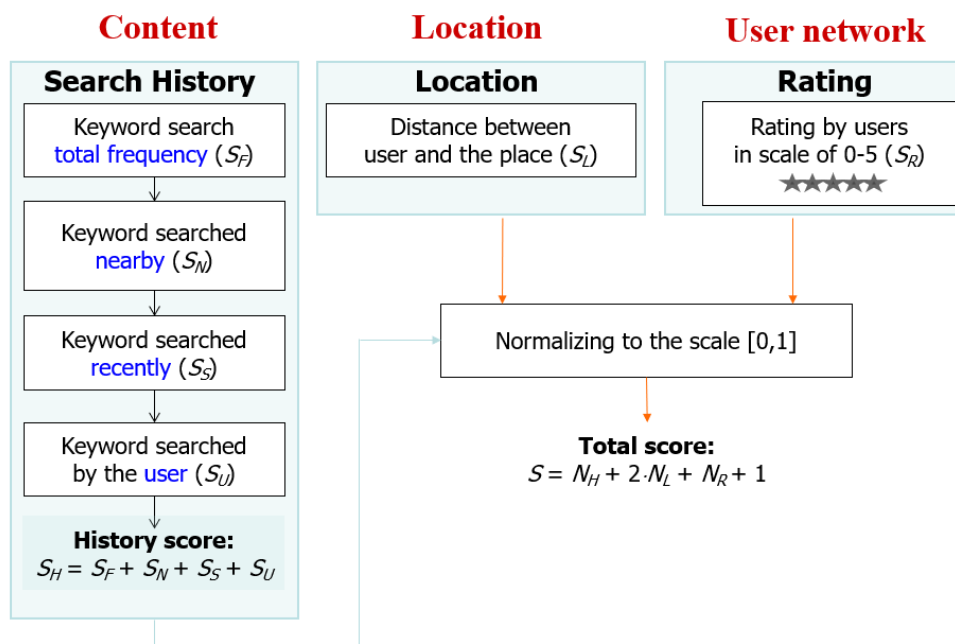
Figure 20. Three factors are used to calculate the relevance of the Mopsi services: (1) content, (2) location, and (3) rating of the service. Content relevance was determined based on the user's search history and by others near the same area. Location relevance was directly the distance. Relevance by the user network was merely the average rating of the service.

6.2. Social network

Mopsi never implemented an explicit social network, although this possibility was always in mind and would have been a relatively straightforward extension. Instead, the platform remained fully open, where all users could connect and see each other's data. In practice, it was a small-scale open social network, as many active users knew each other. Despite its high potential, the social network aspects never turned into a major research direction in Mopsi.

There were two main features of social networks in Mopsi, though. First, there was an operational chat function that allowed users to communicate with each other. This functionality did not use any location-based aspect. The second feature was to share user location. Knowing where the others were traveling and the possibility of seeing their travel photos were the most interesting features for users. Some efforts were made to enhance by predicting where a moving user might go [45]. For example, it would be fun if Mopsi alerted that your friend is about to arrive at your location in 5 minutes.

6.3. Games

Mopsi does not include any games, but there were a few spin-off ideas. The most important is *the Mopsi orienteering* game called *O-Mopsi* [15]. It is a simple treasure hunt game that aims to find a set of real-world targets. The enjoyment comes from exploring the areas, finding the targets, and from the trip planning and navigational challenges. A competitive aspect was included by maintaining ranking lists of the fastest players. The game is also suitable for gamifying sightseeing tours, see Figure 21.
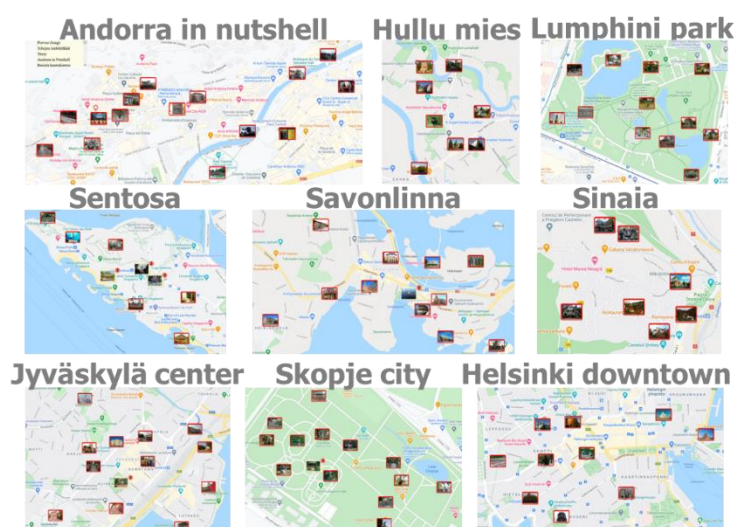
Figure 21. Nine examples of O-Mopsi games created around the world. Typical games appear in parks and other walkable areas in cities.

The design principles of location-based games like O-Mopsi are discussed in depth in [56]. The biggest obstacle of the game was content creation. Finding suitable targets for the games anywhere worldwide is still a big challenge. The most obvious data source is Mopsi, but it lacks coverage. Social media services were considered as they also included a remarkable amount of geotagged content [57,58]. Quality control and suitability of the material is a major challenge in this approach. Open

street map data combined with web mining [59] and web-crawling [60] were also considered, as well as gamifying spatial crowdsourcing [61].

Another game worth mentioning is a location-puzzle game called *Puzzle-Mopsi* [62]. In this game, players can solve the puzzles at home without visiting real-world locations. The game has pictures representing places and a map with equally many locations with empty slots. The goal is to match the photos with the locations. The game used the O-Mopsi database for the game instances and faced the same challenges of having enough content and how to control the quality.

7. Mopsi datasets

Next, we summarize the datasets extracted from Mopsi or collected elsewhere. They were created to test a specific research idea, but other users may invent different uses for them. Most have some ground truth label manually annotated by the researchers or automatically generated by a computer. Some data comes with an interactive webpage demonstrating the research results. The datasets are demonstrated in Figure 22 and summarized in Table 2.
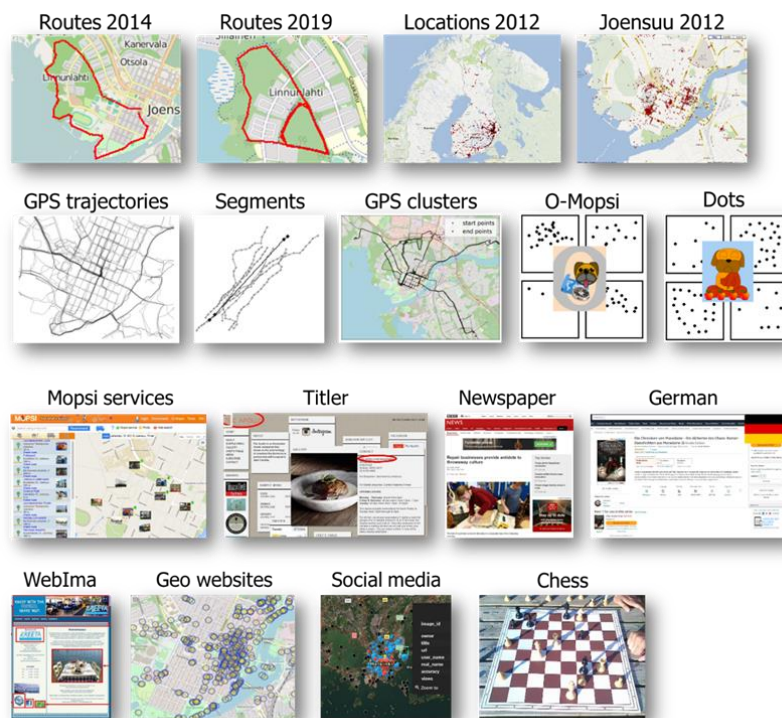

Figure 22. Mopsi datasets include data extracts from the Mopsi collection and additional external data collected mainly for web mining to support Mopsi development.

Table 2. Extracted data subsets from Mopsi[5]. GT denotes ground truth. Reference is the paper where the dataset was introduced.

Dataset:	Ref:	Items:	Collected:	GT:	Description:
Routes 2014	[31]	6,779	2008–2014	User	GPS tracks
Routes 2019	-	2,484	2018–2019	User	GPS tracks
Locations 2012	-	13,467	2008–2012	-	Locations in Finland
Joensuu 2012	-	6,014	2008–2012	-	Locations in Joensuu
Trajectories	[31]	108	2014–2015	Roads	GPS tracks in Joensuu's downtown area
Segments	[32]	355	2014–2015	Roads	Extracted segments from the above
GPS cluster	-	100	2008–2014	Cluster	Trajectories in 10 clusters (Routes 2014)
Chess	-	1050	2009–2019	-	Images with the keyword *Chess*
O-Mopsi	[15]	147	2011–2016	TSP	Set of locations in the O-Mopsi games
Dots	[63]	6449	2017	TSP	Set of points in Dots games
Mopsi services	[48]	414	2010–2017	Keywords	Mopsi services
Title	[48]	1002	2014–2015	Title	Websites with ground truth titles
Newspaper	[50]	2491	2015	Keywords	Newspaper webpages
German	[50]	85	2022	Keywords	Newspaper webpages
WebIma	[51]	1002	2015	Images	Webpages + representative images
Geo websites	[59]	310	2021	Suitability	Webpages + representative images
GeoSoMe	[57]	2027	2019–2022	Suitability	Geotagged images

Routes 2014 is a subset of all routes in Mopsi by that year [31]. It contains 6,779 routes with 7,850,387 points recorded by 51 users. Most routes are in Finland, Joensuu region. *Routes 2019* is another subset of all Mopsi routes created in 2019 but only of the ten most active users and restricted to the Joensuu region. The dataset has 2,484 routes with 3,409,812 points. The user ID can be used as a ground truth.

Locations 2012 contains activity locations in Finland (start or end of tracking and photo taking) by 2012. The data are raw positions without any ground truth. They might be suitable for clustering and for extracting activity areas and patterns. *Joensuu 2012* is the subset containing only points in Joensuu.

Trajectories contain tracks of a single user (Pasi) from 16.11.2014 to 25.4.2015. There were 102 tracks in total, of which only the 45 in the Joensuu area were selected by cropping the data to a square region covering most of downtown [31]. Ground truth road segments and intersections were extracted from the Open Street Map. The data include 355 road segments and 1222 tracks aligned with the segments. The data was used to study road network extraction.

Segments are a processed subset of the trajectories [32]. It includes the ground truth road segments and their intersections. The trajectories were manually assigned with the road segment they belong to (path between two intersections). The data was used to test segment averaging. In addition to the Joensuu segments, Chicago data [11] was also included on the web page.

GPS clusters contain 100 tracks (52,937 points) divided manually into 10 clusters created so that each cluster has tracks having the same start and end location. The cluster ID serves as a ground truth.

The Mopsi Chess dataset was collected by selecting dates of known chess events and dates with

any photo with the keyword *chess*. All photos from these dates were then included. A ground label was assigned to every photo, depending on whether it was related to chess playing. The data can be used to predict this label.

O-Mopsi includes all O-Mopsi game instances [31]. It consists of sets of real-world locations varying in size (4–21) and layout. *Dots* are randomly generated sets of points [63] with random layouts and varying sizes (5–31). The points were designed to have a minimum spread between each other. These datasets have been used for various traveling salesman problems (TSP). Ground truth includes the optimal TSP path (open-loop) and its length.

Mopsi Services [48] are 414 manually created entries in a location-based service database. These include business venues like shops, landmarks, and other points of interest. The entries have manually annotated ground truth titles and keywords. A web link and its content are also stored to allow its use for web mining research.

Titler dataset [48] has 1002 unique web pages collected from Google searches. The data includes original web address, downloaded web content, and a manually annotated ground truth title.

Newspaper [50] has 2470 web pages collected from four English-language newspapers (1250) and six Finnish-language newspapers (1220) by Florian Berger. The idea is that newspaper webpages have annotated ground truth keywords in their metatags in a standardized format. This makes it useful for studies like keyword extraction and topic summarization. *The German* dataset is a similar collection made from German-language newspapers.

The WebIma dataset [51] was created by 117 volunteers who submitted 1002 webpages and up to three images that they consider representative of this website. The image IDs serve as the ground truth. The set was used to study the extraction of a representative image from the web pages.

GeoWebsites is a selection of 330 web pages in OSM [59]. The goal was to study how many of the webpages have representative images qualified to be used as a game target in O-Mopsi. The ground truth is whether the binary label is *suitable* or *not*.

GeoSoMe includes 2027 images taken from Flickr, Yelp, and Google Places around six selected touristic places, such as Stonehenge and Koli National Park [57]. The images were manually inspected for their location accuracy and representativeness in gaming.

8. Conclusions

Many data collection and social media platforms also record the location of data. However, Mopsi is a rare platform that entirely focuses on location. In this paper, we have introduced the main features of the Mopsi platform and summarized the most important research results achieved. In recent years, Mopsi has only had limited maintenance; therefore, many functionalities may not work anymore to demonstrate all its features thoroughly.

We also summarized several extracted datasets that are annotated and freely available for research purposes. We have found them very useful despite of only small-scale. It may not be enough for training machine learning models, but it is highly suitable for testing models trained by other materials. It is real-world, heterogeneous data useful for many location-based research problems. Feel free to contact the author if you have any issues with using the data.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools for this article.

Conflict of interest

The author declares no conflict of interest.

References

1. G. Hariharan, P. Fränti, S. Mehta, Data mining for personal navigation, *Proceedings of Data Mining and Knowledge Discovery: Theory, Tools, and Technology IV*, 2002, 355–365. https://doi.org/10.1117/12.460246

2. A. Tabarcea, N. Gali, P. Fränti, Framework for location-aware search engine, *J. Locat. Based Serv.*, **11** (2017), 50–74. https://doi.org/10.1080/17489725.2017.1407001

3. E. Cho, S. A. Myers, J. Leskovec, Friendship and mobility: user movement in location-based social networks, *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, 1082–1090. https://doi.org/10.1145/2020408.2020579

4. H. Huang, Y. Cheng, W. Dong, G. Gartner, J. M. Krisp, L. Meng, Context modeling and processing in location based services: research challenges and opportunities, *J. Locat. Based Serv.*, **18** (2024), 381–407. https://doi.org/10.1080/17489725.2024.2306349

5. X. Wei, Y. Qian, C. Sun, J. Sun, Y. Liu, A survey of location-based social networks: problems, methods, and future research directions, *GeoInformatica*, **26** (2022), 159–199. https://doi.org/10.1007/s10707-021-00450-1

6. H. Katsumi, W. Yamada, K. Ochiai, Characterizing generic POI: a novel approach for discovering tourist attractions, *Journal of Information Processing*, **31** (2023), 265–277. https://doi.org/10.2197/ipsjjip.31.265

7. K. Waga, A. Tabarcea, P. Fränti, Context aware recommendation of location-based data, *Proceedings of 15th International Conference on System Theory, Control and Computing*, 2011, 658–663.

8. K. Waga, A. Tabarcea, P. Fränti, Recommendation of points of interest from user generated data collection, *Proceedings of 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 2012, 550–555. https://doi.org/10.4108/icst.collaboratecom.2012.250451

9. P. Fränti, J. Chen, A. Tabarcea, Four aspects of relevance in location-based media: content, time, location and network, *Proceedings of the 7th International Conference on Web Information Systems and Technologies*, 2011, 413–417.

10. R. Mariescu-Istodor, P. Fränti, CellNet: inferring road networks from GPS trajectories, *ACM Trans. Spat. Algor.*, **4** (2018), 8. https://doi.org/10.1145/3234692

11. J. Biagioni, J. Eriksson, Inferring road maps from global positioning system traces: survey and comparative evaluation, *Transport. Res. Rec.*, **2291** (2012), 61–71. https://doi.org/10.3141/2291-08

12. M. Ahmed, S. Karagiorgou, D. Pfoser, C. Wenk, A comparison and evaluation of map construction algorithms using vehicle tracking data, *GeoInformatica*, **19** (2015), 601–632. https://doi.org/10.1007/s10707-014-0222-6

13. M. H. Sharker, H. A. Karimi, J. C. Zgibo, Health-optimal routing in pedestrian navigation services, *Proceedings of the First ACM SIGSPATIAL International Workshop on Use of GIS in Public Health*, 2012, 1–10. https://doi.org/10.1145/2452516.2452518

14. M. H. Sharker, H. A. Karimi, Computing least air pollution exposure routes, *Int. J. Geogr. Inf. Sci.*, **28** (2014), 343–362. https://doi.org/10.1080/13658816.2013.841317

15. P. Fränti, R. Mariescu-Istodor, L. Sengupta, O-Mopsi: mobile orienteering game for sightseeing, exercising and education, *ACM Trans. Multim. Comput.*, **13** (2017), 56. https://doi.org/10.1145/3115935

16. B. de Moura Morceli, A. Porfirio Dal Poz, Road extraction from low-cost GNSS-device dense trajectories, *J. Locat. Based Serv.*, **17** (2023), 251–270. https://doi.org/10.1080/17489725.2023.2216670

17. J. Yang, X. Ye, B. Wu, Y. Gu, Z. Wang, D. Xia, et al., DuARE: automatic road extraction with aerial images and trajectory data at Baidu maps, *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, 4321–4331. https://doi.org/10.1145/3534678.3539029

18. J. Tang, M. Deng, J. Huang, H. Liu, X. Chen, An automatic method for detection and update of additive changes in road network with GPS trajectory data, *ISPRS Int. J. Geo-Inf.*, **8** (2019), 411. https://doi.org/10.3390/ijgi8090411

19. R. Mariescu-Istodor, P. Fränti, Detecting location-based user actions, *Proceedings of the 14th International Conference on Location Based Services*, 2018, 1–6. https://doi.org/10.3929/ethz-b-000225579

20. R. Bakhshandeh, M. Samadi, Z. Azimifar, J. Schaeffer, Degrees of separation in social networks archived, *Proceedings of the International Symposium on Combinatorial Search*, 2011, 18–23. https://doi.org/10.1609/socs.v2i1.18200

21. P. Fränti, K. Waga, C. Khurana, Can social network be used for location-aware recommendation? *Proceedings of the 11th International Conference on Web Information Systems and Technologies*, 2015, 558–565. https://doi.org/10.5220/0005495805580565

22. M. Fatemi, K. Kucher, M. Laitinen, P. Fränti, Self-similarity of Twitter users, *Proceedings of Swedish Workshop on Data Science (SweDS)*, 2021, 1–7. https://doi.org/10.1109/SweDS53855.2021.9638288

23. H. Jiang, J. Li, P. Zhao, F. Zeng, Z. Xiao, A. Iyengar, Location privacy-preserving mechanisms in location-based services: a comprehensive survey, *ACM Comput. Surv.*, **54** (2021), 4. https://doi.org/10.1145/3423165

24. P. Fränti, S. Sieranoja, K. Wikström, T. Laatikainen, Clustering diagnoses from 58M patient visits in Finland between 2015 and 2018, *JMIR Med. Inf.*, **10** (2022), e35422. https://doi.org/10.2196/35422

25. P. Fränti, R. Mariescu-Istodor, A. Akram, M. Satokangas, E. Reissell, Can we optimize locations of hospitals by minimizing the number of patients at risk? *BMC Health Serv. Res.*, **23** (2023), 415. https://doi.org/10.1186/s12913-023-09375-x

26. S. Heitmann, A. Pagotto, C. Kray, Approximate vs. precise location in popular location-based services, *J. Locat. Based Serv.*, in press. https://doi.org/10.1080/17489725.2024.2310006

27. P. Fränti, Mobile navigation for a broad consumer's market: dynamic map handling, *GIM Int.*, **17** (2003), 28–31.

28. P. Fränti, A. Tabarcea, J. Kuittinen, V. Hautamäki, Location-based search engine for multimedia phones, *Proceedings of IEEE International Conference on Multimedia and Expo*, 2010, 558–563. https://doi.org/10.1109/ICME.2010.5583538

29. P. Fränti, J. Kuittinen, A. Tabarcea, L. Sakala, MOPSI location-based search engine: concept, architecture and prototype, *Proceedings of the 2010 ACM Symposium on Applied Computing*, 2010, 872–873. https://doi.org/10.1145/1774088.1774268

30. M. Rezaei, P. Fränti, Real-time clustering of large geo-referenced data for visualizing on map, *Adv. Electr. Comput. Eng.*, **18** (2018), 63–74. https://doi.org/10.4316/AECE.2018.04008

31. R. Mariescu-Istodor, P. Fränti, Grid-based method for GPS route analysis for retrieval, *ACM Trans. Spat. Algor.*, **3** (2017), 8. https://doi.org/10.1145/3125634

32. P. Fränti, R. Mariescu-Istodor, Averaging GPS segment competition 2019, *Pattern Recogn.*, **112** (2021), 107730. https://doi.org/10.1016/j.patcog.2020.107730

33. K. Waga, A. Tabarcea, R. Mariescu-Istodor, P. Fränti, Real time access to multiple GPS tracks, *Proceedings of the 9th International Conference on Web Information Systems and Technologies*, 2013, 293–299. https://doi.org/10.5220/0004370102930299

34. M. Chen, M. Xu, P. Fränti, A fast O(N) multi-resolution polygonal approximation algorithm for GPS trajectory simplification, *IEEE Trans. Image Process.*, **21** (2012), 2770–2785. https://doi.org/10.1109/TIP.2012.2186146

35. R. Mariescu-Istodor, P. Fränti, Context-aware similarity of GPS trajectories, *J. Locat. Based Serv.*, **14** (2020), 231–251. https://doi.org/10.1080/17489725.2020.1842923

36. R. Mariescu-Istodor, P. Fränti, Gesture input for GPS route search, In: *Structural, syntactic, and statistical pattern recognition*, Cham: Springer, 2016, 439–449. https://doi.org/10.1007/978-3-319-49055-7_39

37. K. Waga, A. Tabarcea, M. Chen, P. Fränti, Detecting movement type by route segmentation and classification, *Proceedings of 8th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*, 2012, 1–6. https://doi.org/10.4108/icst.collaboratecom.2012.250450

38. R. Mariescu-Istodor, A. Tabarcea, R. Saeidi, P. Fränti, Low complexity spatial similarity measure of GPS trajectories, *Proceedings of the 10th International Conference on Web Information Systems and Technologies (WEBIST-2014)*, 2014, 62–69. https://doi.org/10.5220/0004940500620069

39. R. Mariescu-Istodor, P. Heng, P. Fränti, Roundness measure for GPS routes, *Proceedings of the 14th International Conference on Location Based Services*, 2018, 81–86. https://doi.org/10.3929/ethz-b-000225594

40. J. W. Yang, R. Mariescu-Istodor, P. Fränti, Three rapid methods for averaging GPS segments, *Appl. Sci.*, **9** (2019), 4899. https://doi.org/10.3390/app9224899

41. B. Jimoh, R. Mariescu-Istodor, P. Fränti, Is medoid suitable for averaging GPS trajectories? *ISPRS Int. J. Geo-Inf.*, **11** (2022), 133. https://doi.org/10.3390/ijgi11020133

42. R. Mariescu-Istodor, P. Fränti, Fast travel distance estimation using overhead graph, *J. Locat. Based Serv.*, **15** (2021), 261–279. https://doi.org/10.1080/17489725.2021.1889058

43. S. Sieranoja, T. Kinnunen, P. Fränti, GPS trajectory biometrics: from where you were to how you move, In: *Structural, syntactic, and statistical pattern recognition*, Cham: Springer, 2016, 450–460. https://doi.org/10.1007/978-3-319-49055-7_40

44. P. Fränti, R. Mariescu-Istodor, K. Waga, Similarity of mobile users based on sparse location history, In: *Artificial intelligence and soft computing*, Cham: Springer, 2018, 593–603. https://doi.org/10.1007/978-3-319-91253-0_55

45. R. Mariescu-Istodor, R. Ungureanu, P. Fränti, Real-time destination prediction for mobile users, *Adv. Cartogr. GIScience Int. Cartogr. Assoc.*, **2** (2019), 1–7. https://doi.org/10.5194/ica-adv-2-10-2019

46. P. Fränti, Mopsi routes: creative ways to analyze GPS tracks, manuscript.

47. N. Gali, P. Fränti, Content-based title extraction from web page, *Proceedings of the 12th International Conference on Web Information Systems and Technologies*, 2016, 204–210. https://doi.org/10.5220/0005794102040210

48. N. Gali, R. Mariescu-Istodor, P. Fränti, Using linguistic features to automatically extract web page title, *Expert Syst. Appl.*, **79** (2017), 296–312. https://doi.org/10.1016/j.eswa.2017.02.045

49. M. Rezaei, N. Gali, P. Fränti, ClRank: a method for keyword extraction from web pages using clustering and distribution of nouns, *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 2015, 79–84. https://doi.org/10.1109/WI-IAT.2015.64

50. H. Shah, P. Fränti, Combining statistical, structural, and linguistic features for keyword extraction from web pages, *Appl. Comput. Intell.*, **2** (2022), 115–132. https://doi.org/10.3934/aci.2022007

51. N. Gali, A. Tabarcea, P. Fränti, Extracting representative image from web page, *Proceedings of the 11th International Conference on Web Information Systems and Technologies*, 2015, 411–419. https://doi.org/10.5220/0005438704110419

52. N. Gali, R. Mariescu-Istodor, P. Fränti, Functional classification of websites, *Proceedings of the 8th International Symposium on Information and Communication Technology*, 2017, 34–41. https://doi.org/10.1145/3155133.3155178

53. N. Gali, R. Mariescu-Istodor, P. Fränti, Similarity measures for title matching, *Proceedings of 23rd International Conference on Pattern Recognition (ICPR)*, 2016, 1549–1554. https://doi.org/10.1109/ICPR.2016.7899857

54. N. Gali, R. Mariescu-Istodor, D. Hostettler, P. Fränti, Framework for syntactic string similarity measures, *Expert Syst. Appl.*, **129** (2019), 169–185. https://doi.org/10.1016/j.eswa.2019.03.048

55. R. Mariescu-Istodor, A. S. M. Sayem, P. Fränti, Activity event recommendation and attendance prediction, *J. Locat. Based Serv.*, **13** (2019), 293–319. https://doi.org/10.1080/17489725.2019.1660423

56. P. Fränti, N. Fazal, Design principles for content creation in location-based games, *ACM Trans. Multim. Comput.*, **19** (2023), 165. https://doi.org/10.1145/3583689

57. N. Fazal, P. Fränti, Social media data for content creation in location-based games, *J. Locat. Based Serv.*, in press. https://doi.org/10.1080/17489725.2024.2414000

58. N. Fazal, P. Fränti, Relevant tag extraction based on image visual content, In: *Applied intelligence*, Singapore: Springer, 2024, 283–295. https://doi.org/10.1007/978-981-97-0827-7_25

59. N. Fazal, R. Mariescu-Istodor, P. Fränti, Using open street map for content creation in location-based games, *Proceedings of 29th Conference of Open Innovations Association (FRUCT)*, 2021, 109–117. https://doi.org/10.23919/FRUCT52173.2021.9435502

60. N. Fazal, K. Q. Nqyuen, P. Fränti, Efficiency of web crawling for geo-tagged image retrieval, *Webology*, **16** (2019), 16–39. https://doi.org/10.14704/WEB/V16I1/A177

61. N. Fazal, P. Fränti, Mopsify: gamified crowdsourcing for content creation in location-based games, *Proceedings of the Sixteenth International Conference on Advanced Geographic Information Systems, Applications, and Services*, 2024, 18–22.

62. P. Fränti, L. Kong, Puzzle-Mopsi: a location-puzzle game, *Appl. Comput. Intell.*, **3** (2023), 1–12. https://doi.org/10.3934/aci.2023001

63. L. Sengupta, R. Mariescu-Istodor, P. Fränti, Planning your route: where to start? *Comput. Brain Behav.*, **1** (2018), 252–265. https://doi.org/10.1007/s42113-018-0018-0