UNIVERSITY OF
EASTERN FINLAND

Web tools for analyzing location-based data

Pasi Fränti

3rd December, 2021

Plenary talk:

12th Conference on Data Analysis Methods for Software Systems

University of Eastern Finland

UNIVERSITY OF EASTERN FINLAND

2001

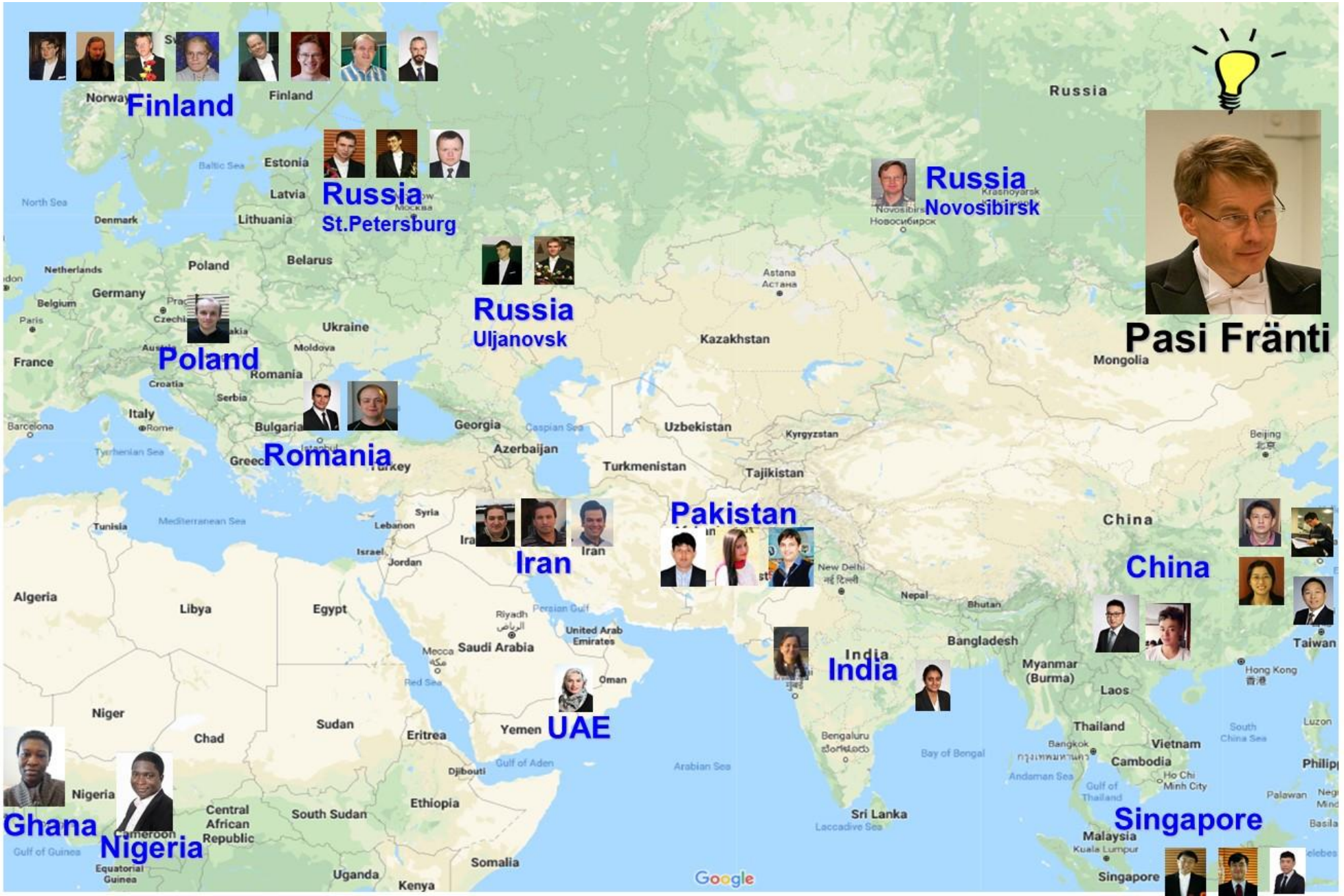
2013

2018

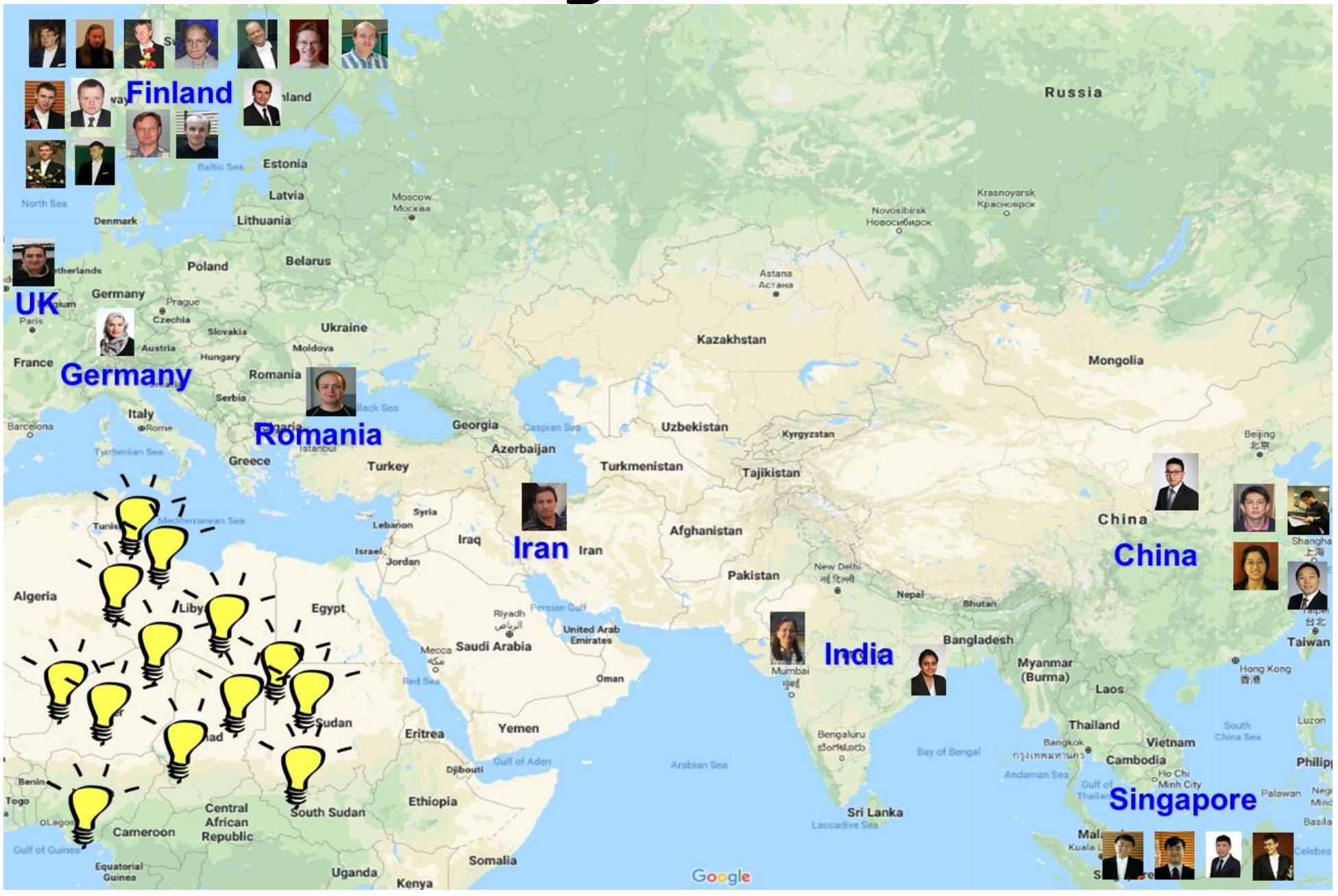
2017

2020

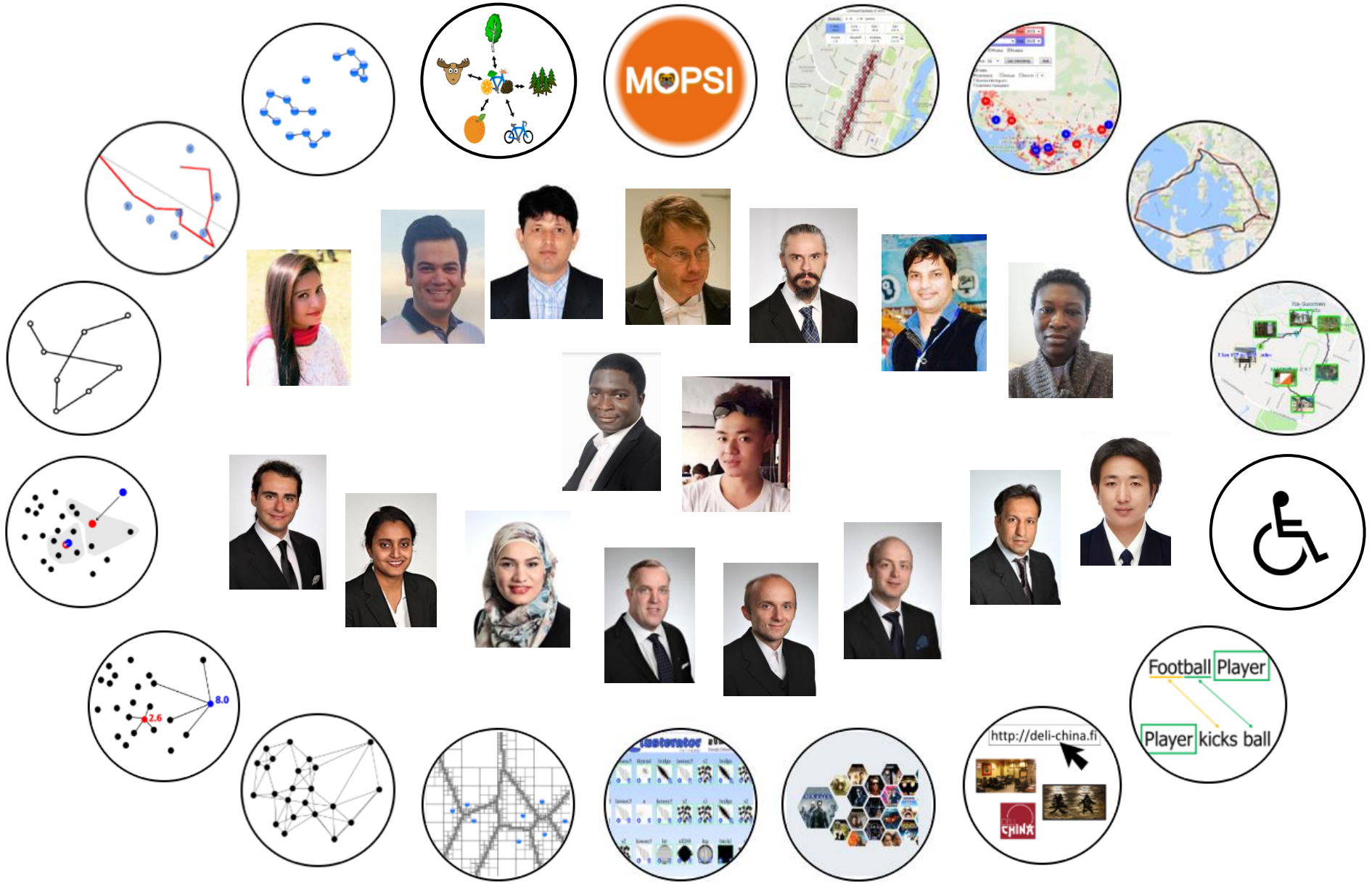
Start locations



After graduation



Machine learning group



Research

Images

Compression and data reduction
Segmentation
Denoising
HDR

Location-based

Storage and retrieval
Search and recommendation
Route analysis
Games

Clustering methods

Algorithms
Outliers
Validity
Graphs
Strings
Sets
Categories

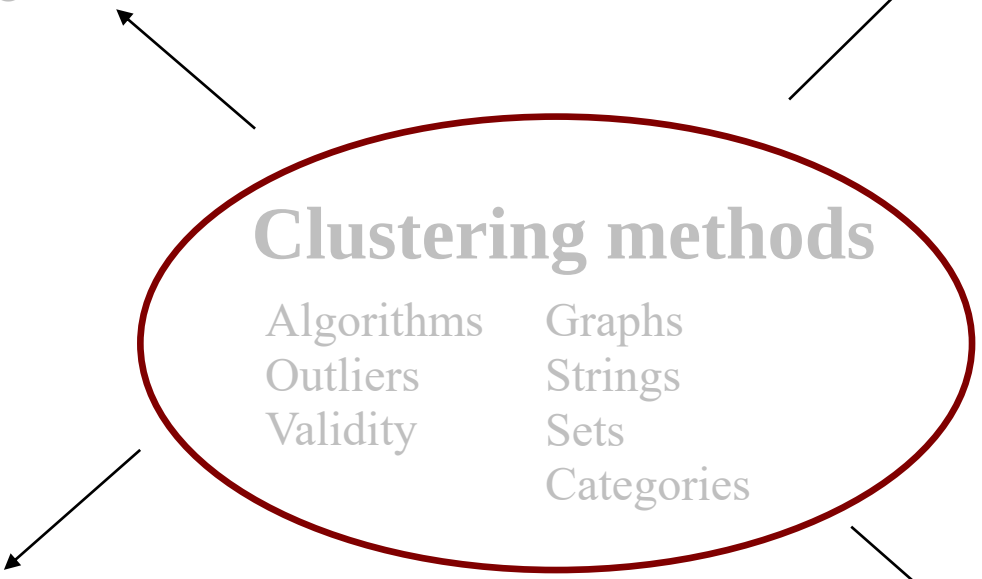
Voice

Speaker recognition
Voice activity detection
Applications

Health care

Service location optimization
Multimorbidity patterns
Disease prediction
Heart rate variability

Focus today



Clustering methods

Algorithms
Outliers
Validity
Graphs
Strings
Sets
Categories

Images

Compression and data reduction
Segmentation
Denoising
HDR

Location-based

Storage and retrieval
Search and recommendation
Route analysis
Games

Voice

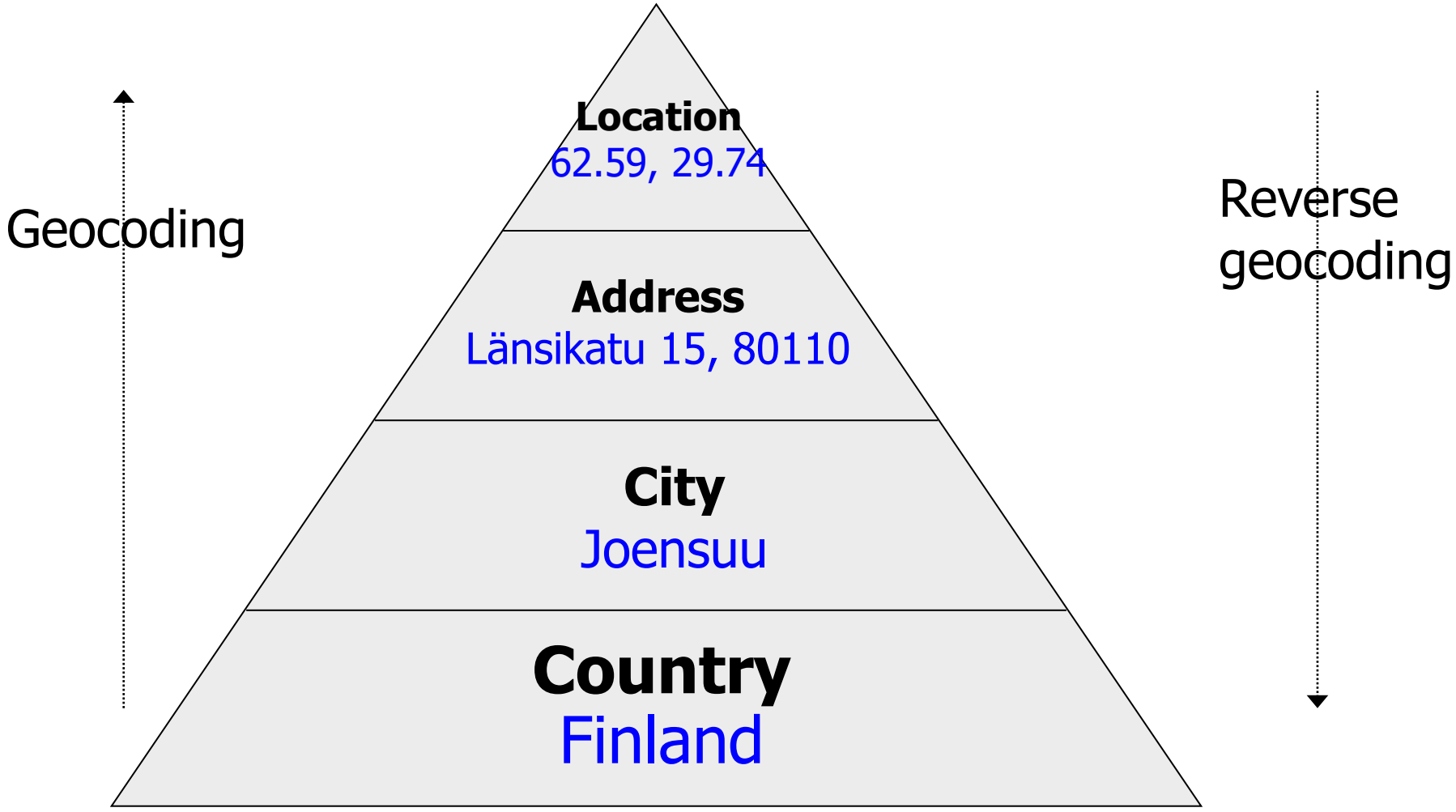
Speaker recognition
Voice activity detection
Applications

Health care

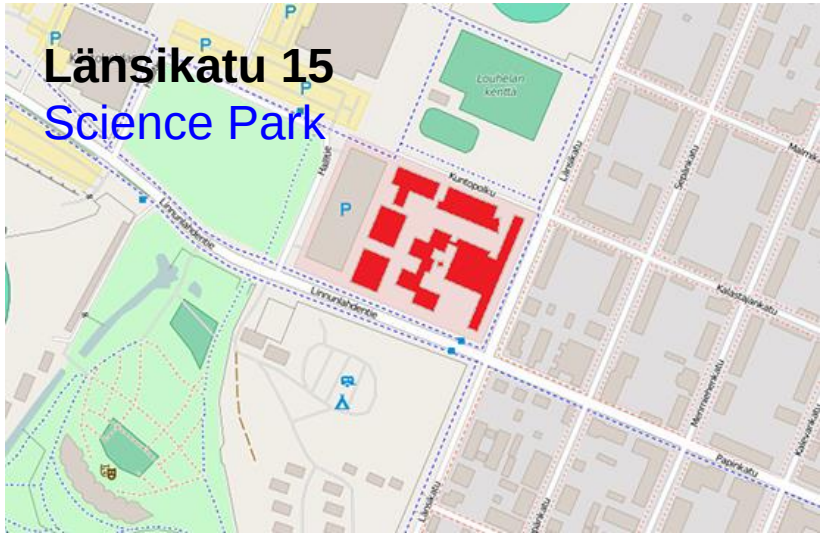
Service location optimization
Multimorbidity patterns
Disease prediction
Heart rate variability

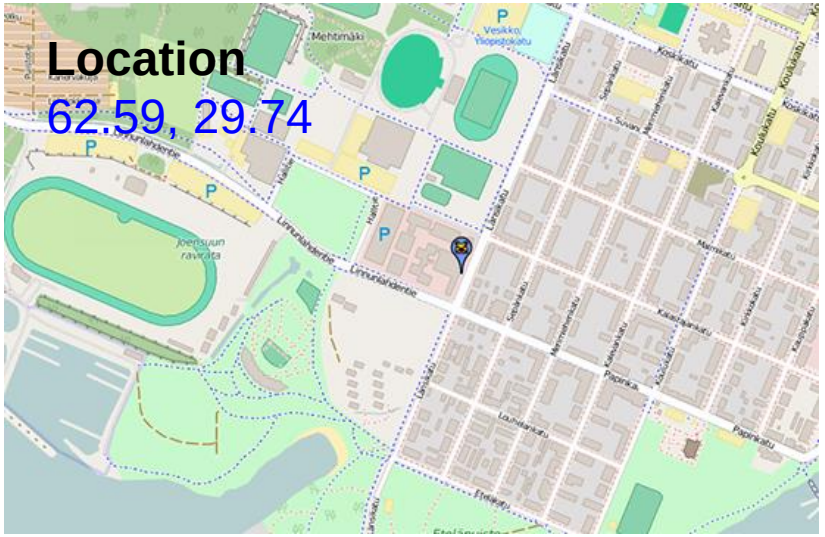
Location

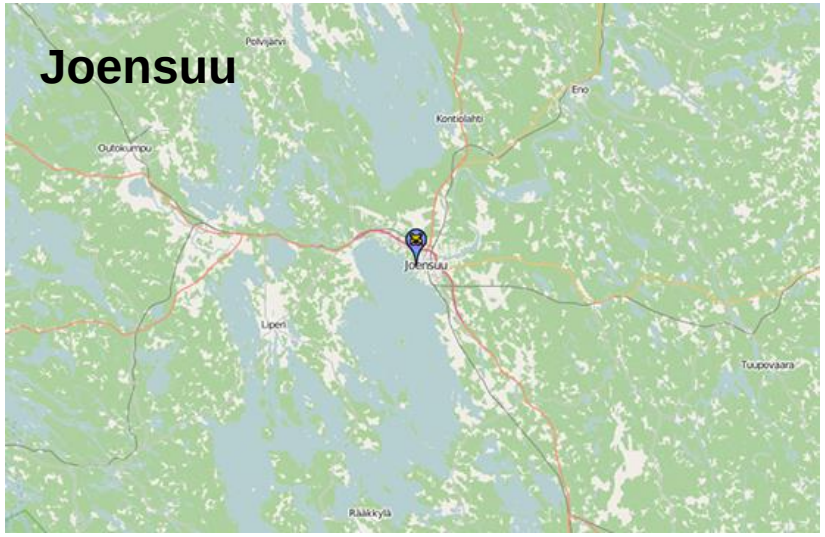
Location hierarchy



Levels of location







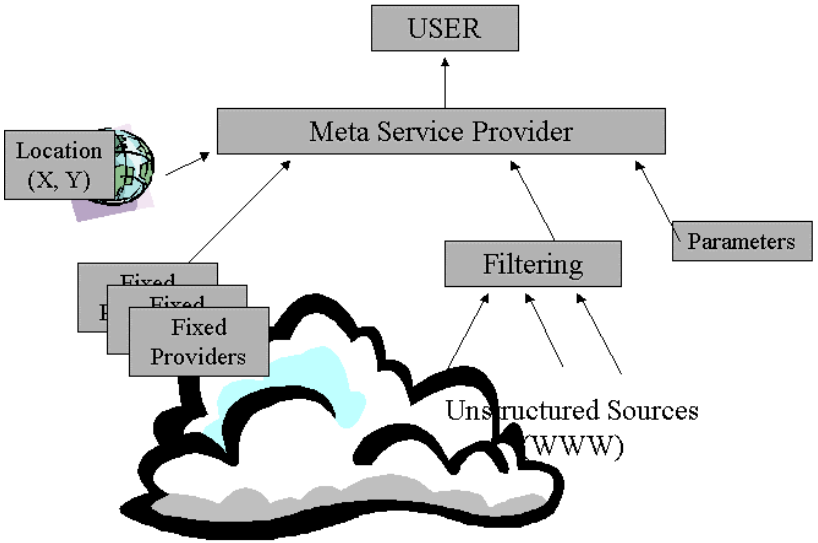
Mopsi

http://cs.uef.fi/mopsi/

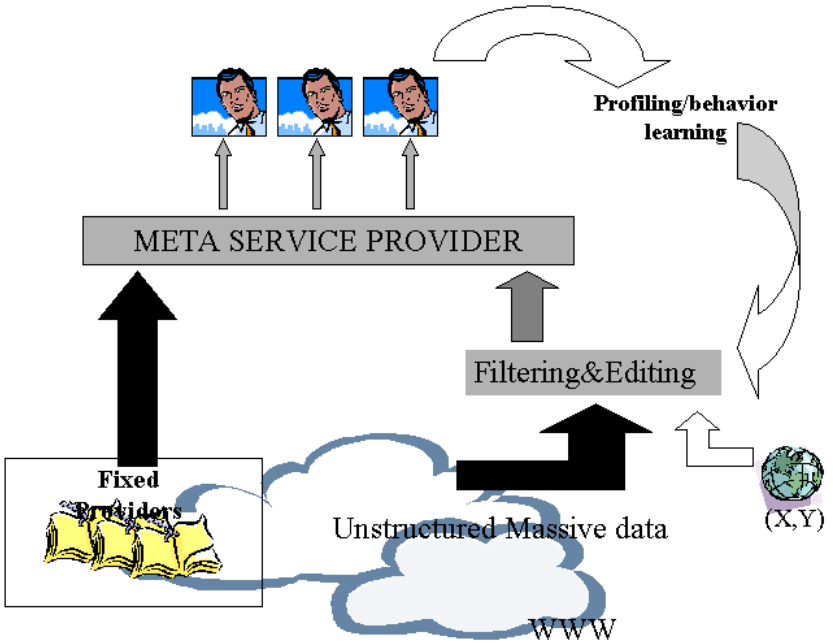
2002
The very first idea


Hariharan, Fränti and Mehta
Data Mining for Personal Navigation
Data Mining and Knowledge Discovery: Theory, Tools, and Technology IV, **2002**

Location-based search



User-profiling added



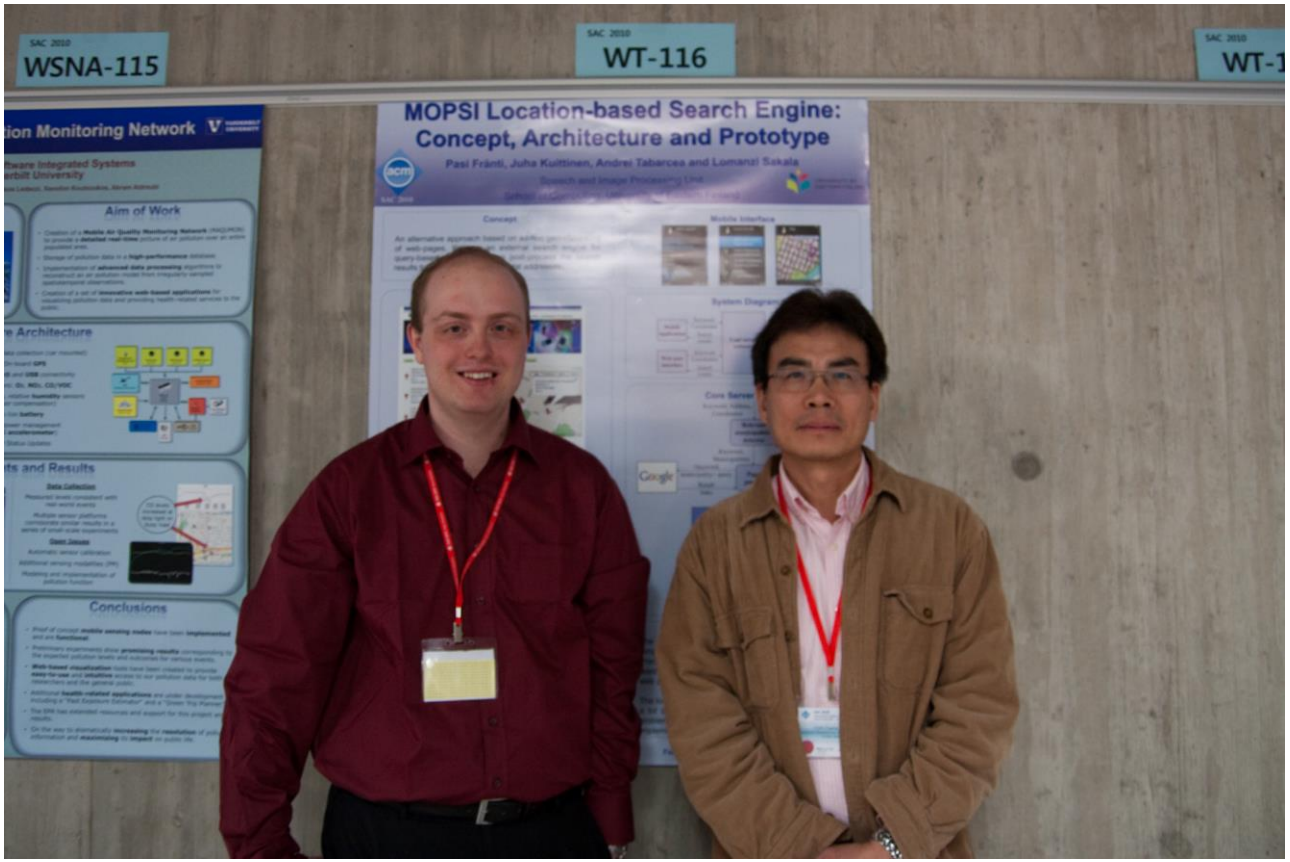
GoogleMaps was launched **2005**

2010
Mopsi first prototype

Fränti, Kuittinen, Tabarcea, Sakala, "MOPSI location-based search engine: concept, architecture and prototype", *ACM SAC,* **2010**

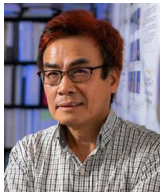
2006: Working prototype (**50%** relevance)

2009: First (meta) LBS search-engine



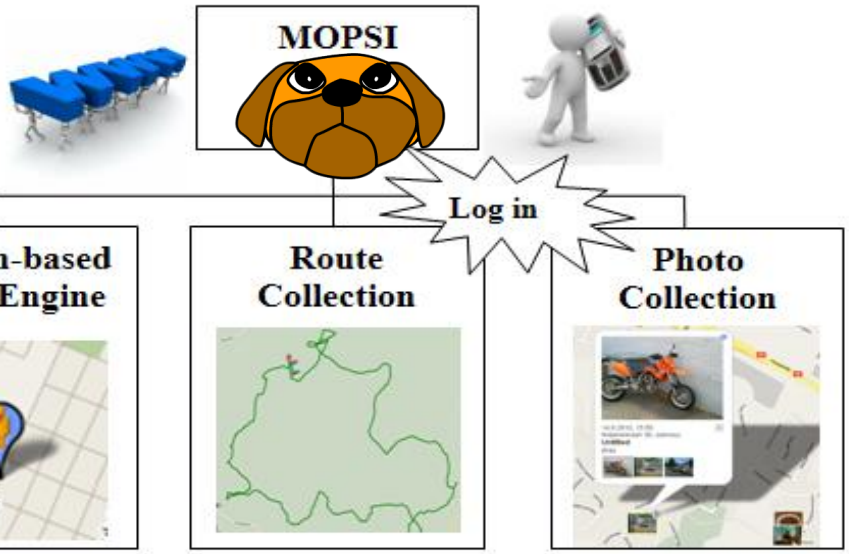
J. Kuittinen

A. Tabarcea



Chih-Cheng Hung

2011
Mopsi overview



Q. Zhao

Recommendation System

- Service (bus, friend)
- Text (search query, photo description)

Web Content Mining

- Meta searching
- Service title detection
- Document processing

Route Pattern

- Route reduction
- Route segmentation
- Activity area

Social network

- Facebook

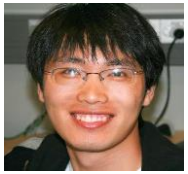
Location-based Game

- Orienteering
- Killer-game

Data collection in Mopsi

First photos: 2009

Other users:




www

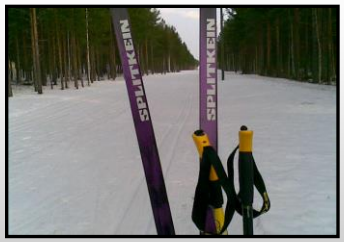
MOPSI webpage

Service directory

Data collector:


GPS

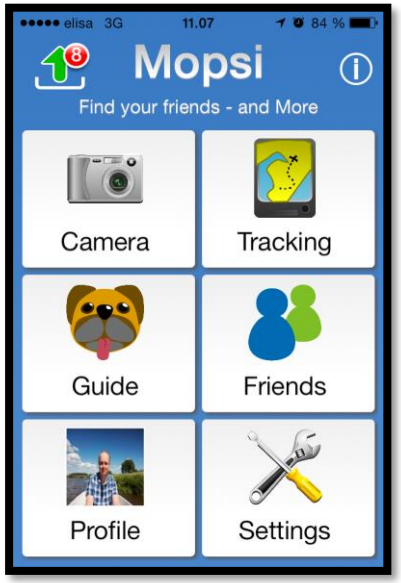
User: Pasi

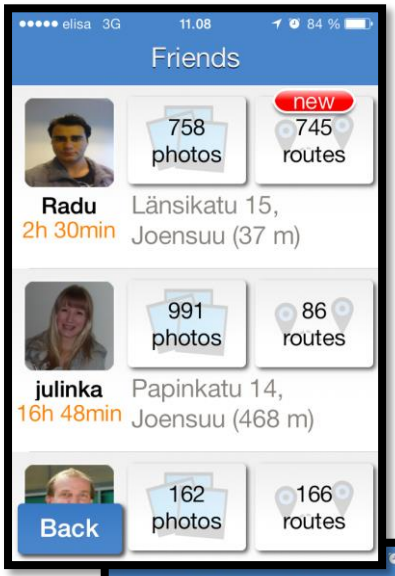


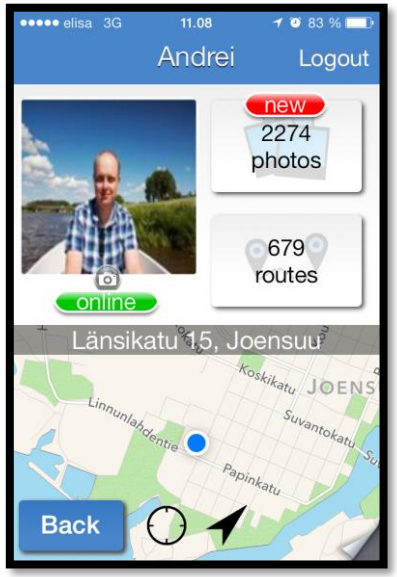
N 62.63 E 29.86
Last skiing of winter

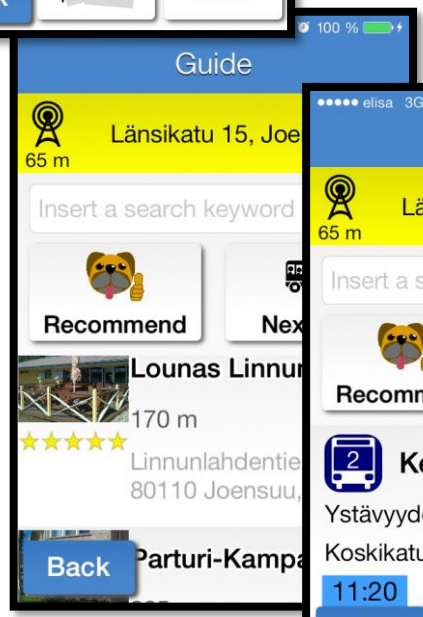
User collection

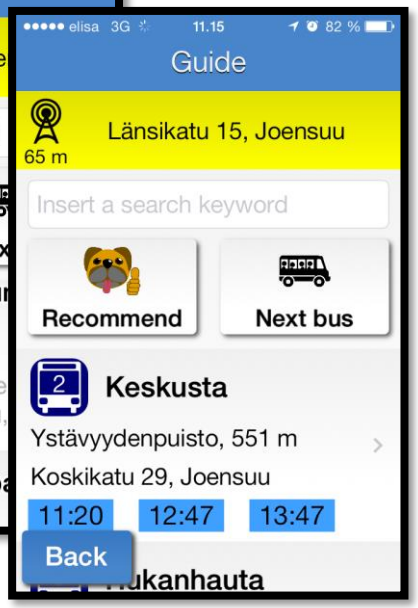
Mobile Application









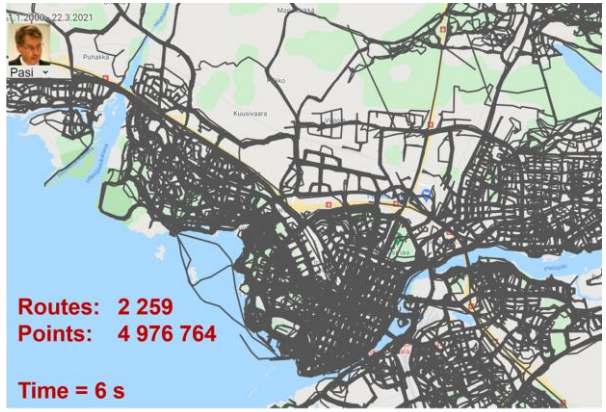


Data collected

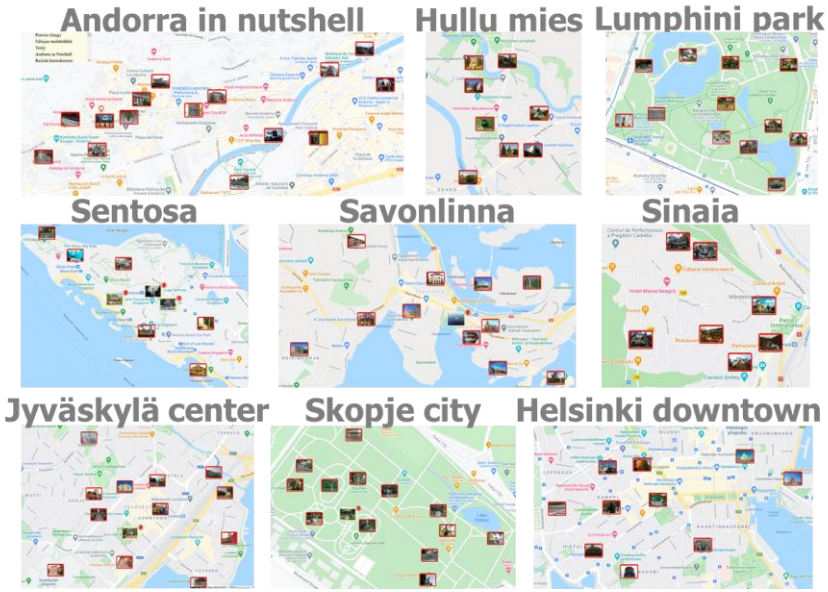
Data collected

27.2.2009 – 14.9.2020

Users:	**203**
Photos:	**64,346**
Tracks:	**11,449**
Services:	**414**
Games:	**158**



5310	Partiomaja - Youth Hostel			rezaei	maa	1.1.2010	3.3.2016	Confirmed	retkeilymaja, hostel
5309	Lykynlampi sauna			Pasi	Pasi	1.1.2010	5.1.2016	Confirmed	sauna
5307	Joensuun kirkko			Oili	Pasi	1.1.2010	22.4.2015	Confirmed	kirkko, church
5303	Utran kirkko			Oili		1.1.2010	1.1.2010	Confirmed	church, kirkko
5301	Helluntaikirkko			Oili	Pasi	1.1.2010	5.1.2016	Confirmed	kirkko, church, helluntaikirkko
5300	K-market Suvantokatu			Wan	Pasi	1.1.2010	20.9.2019	Confirmed	kauppa, valintatalo, ruoka
5299	Mokkamaa			rezaei	rezaei	1.1.2010	12.2.2014	Confirmed	kahvi, tee, suklaa, herkkupuoti, erikoiskahvi
5298	Dressmann			Pasi	Pasi	1.1.2010	25.2.2014	Confirmed	vaatekauppa
5297	KappAhl			Pasi	Pasi	1.1.2010	25.2.2014	Confirmed	vaatekauppa, muoti

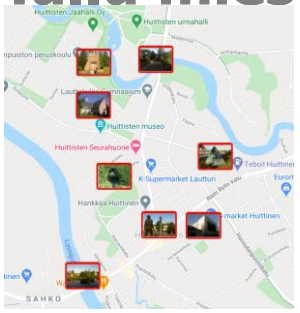


O-Mopsi games

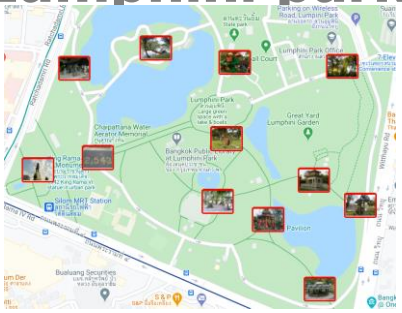
10 selected examples

Andorra in nutshell

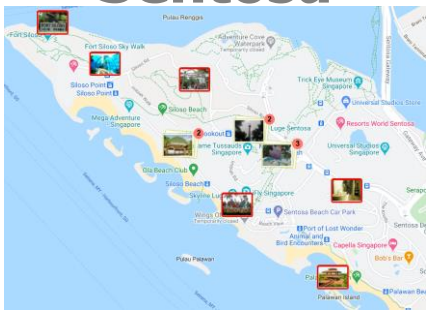
Hullu mies



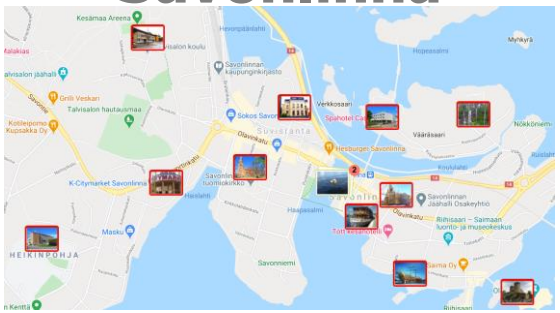
Lumphini park



Sentosa

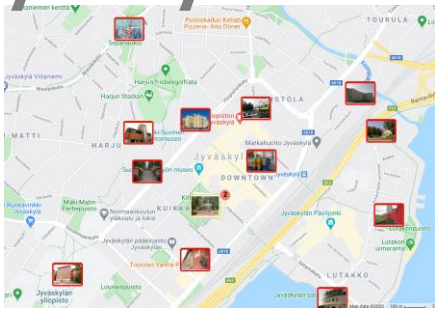


Savonlinna

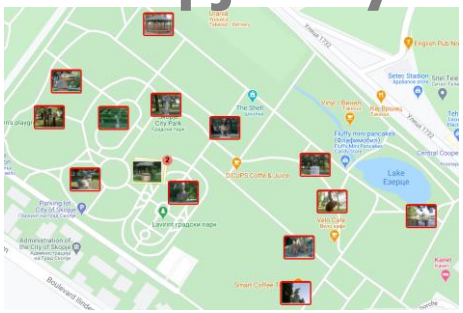


Sinaia

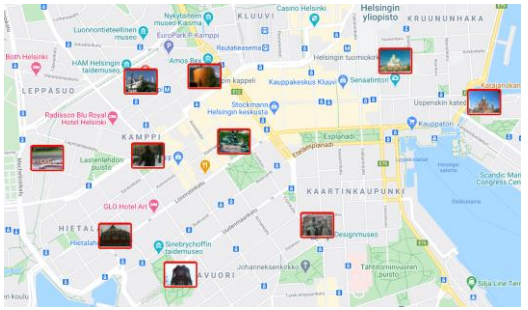
Jyväskylä center



Skopje city

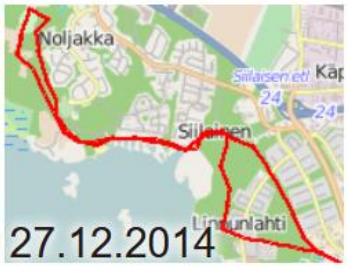


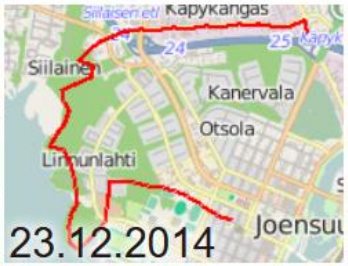
Helsinki downtown

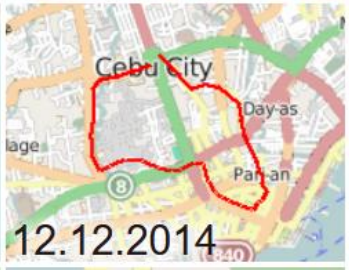


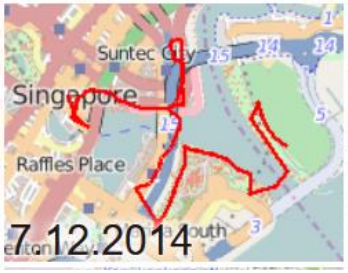
Mopsi Routes 2014

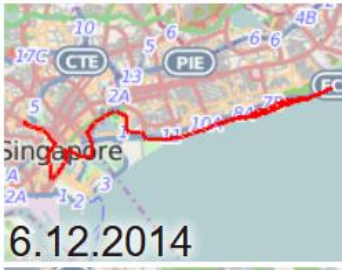
12 selected examples











Summary of data:

Routes	6,779
Points	7,850,387
Kilometers	87,851
Hours	4,504

Collection details:

Who	51 Mopsi Users
When	19.7.2008 - 31.12.2014
How	Various movement types
Issues	Some GPS errors are present

Web-applications

1. Photos
2. GPS tracks
3. Games and others

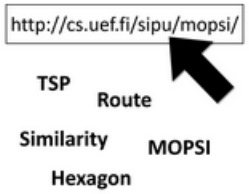
Selected tools

http://cs.uef.fi/sipu/mopsi/

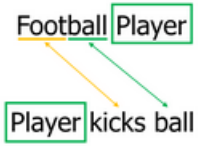
Web and text

Image Extraction

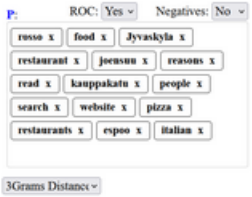
Keyword Extraction (clRank)



String Similarity



Soft precision & recall



Routes

Overhead graph

Road network

Route Similarity

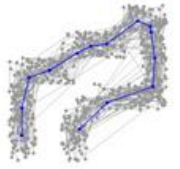
Context-aware

Transport Mode

Polygonal Approx.

Route Dataset

Segment averagi



Clustering

Marker clustering

Cluster creator

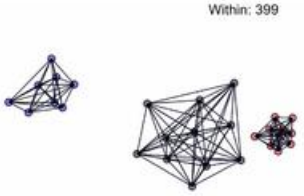
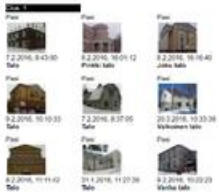


Photo descriptions



Other

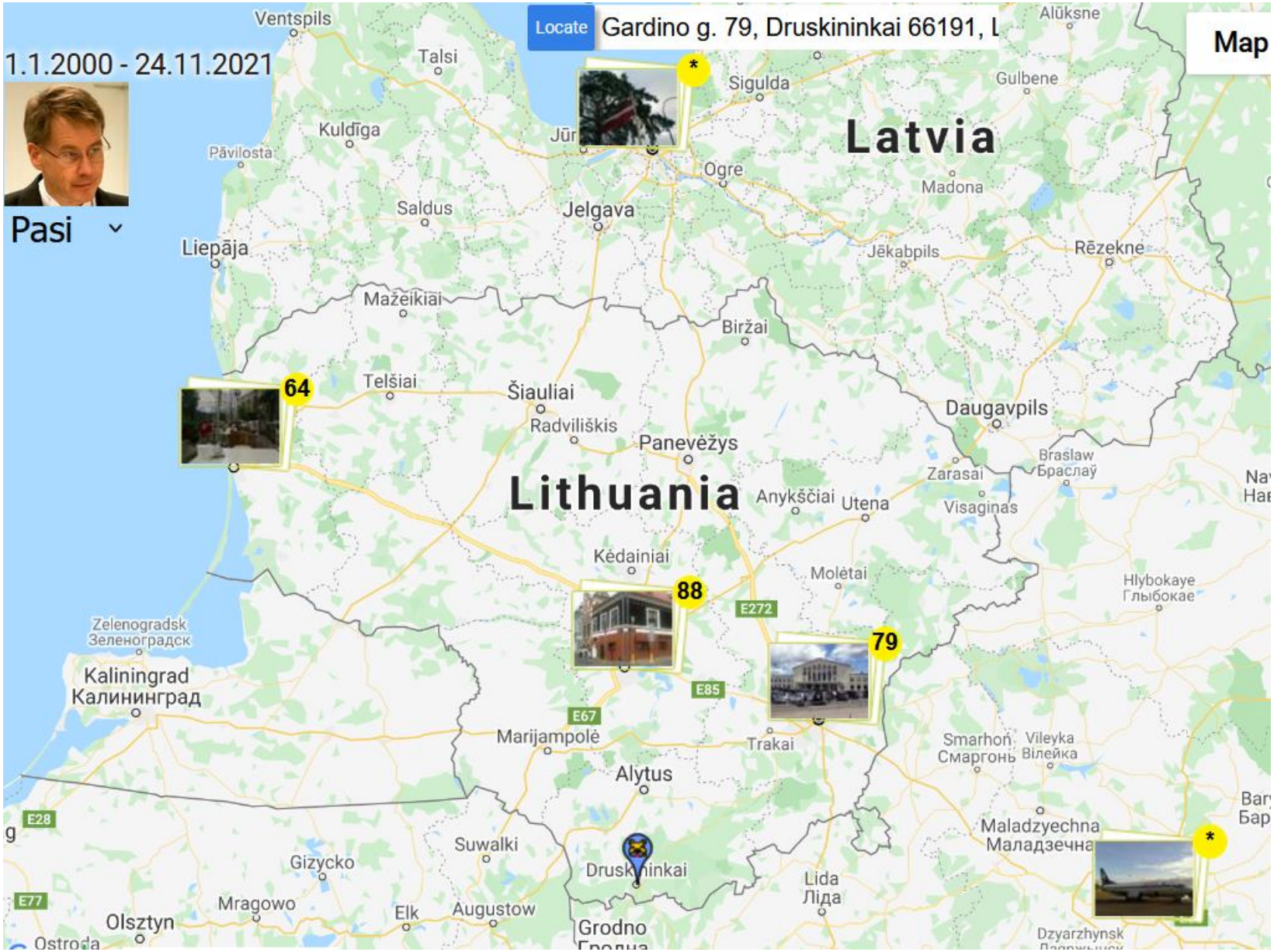
XNN

User Similarity

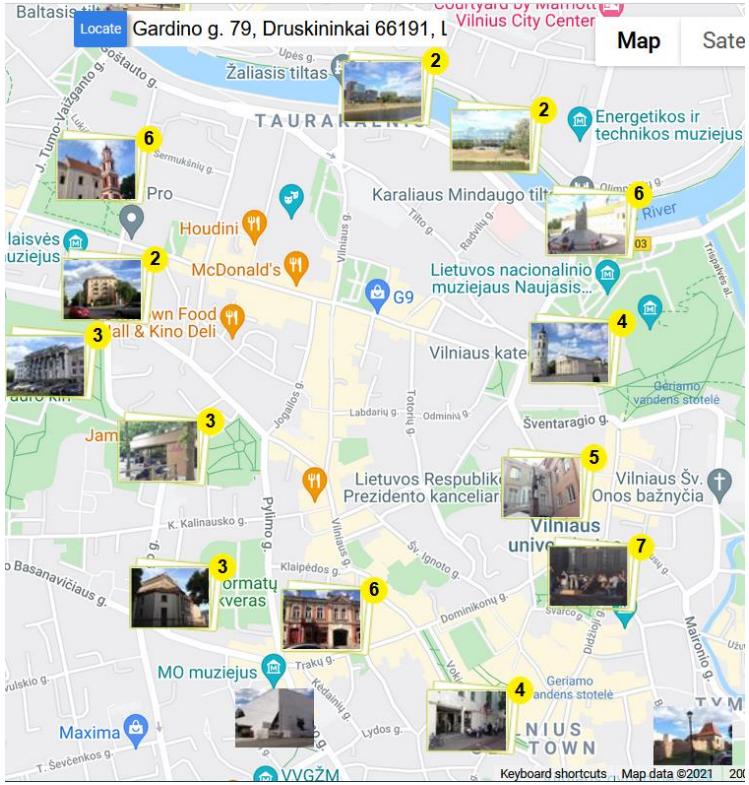
Mopsi events

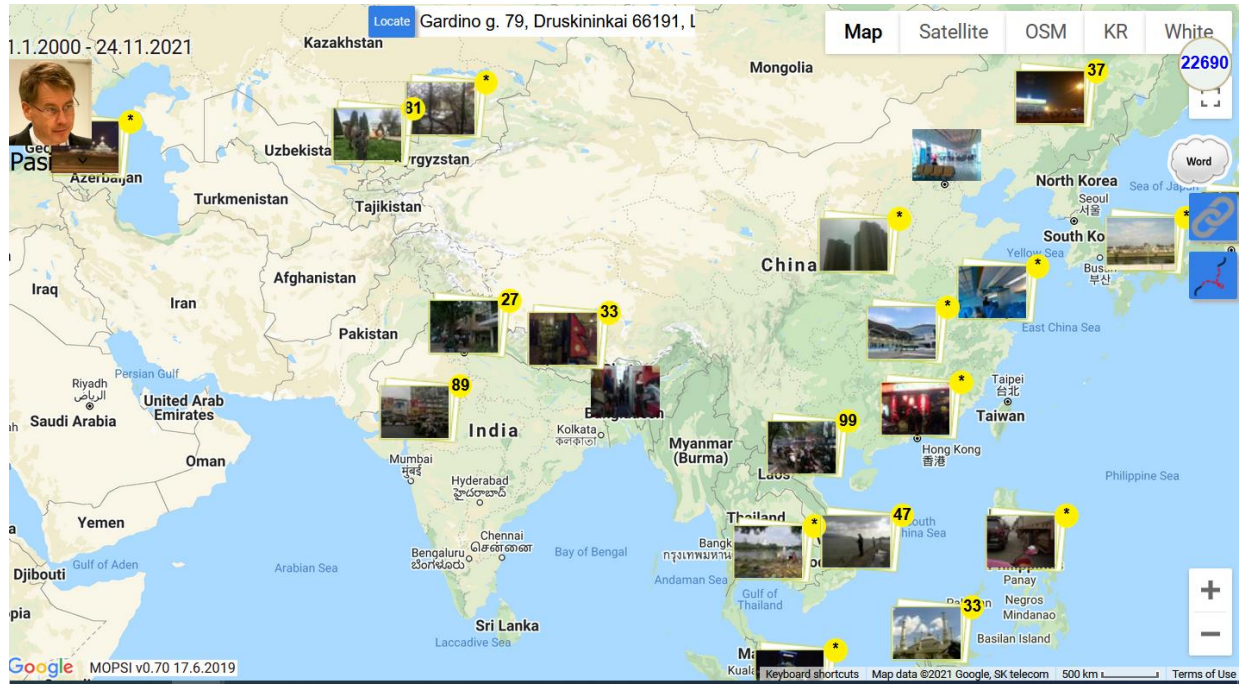
Photos

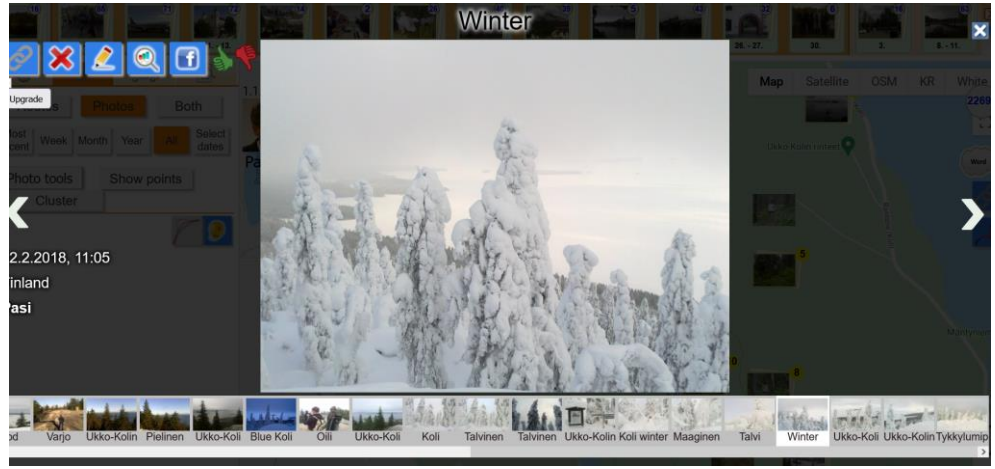
Selected locations on map



Selected views

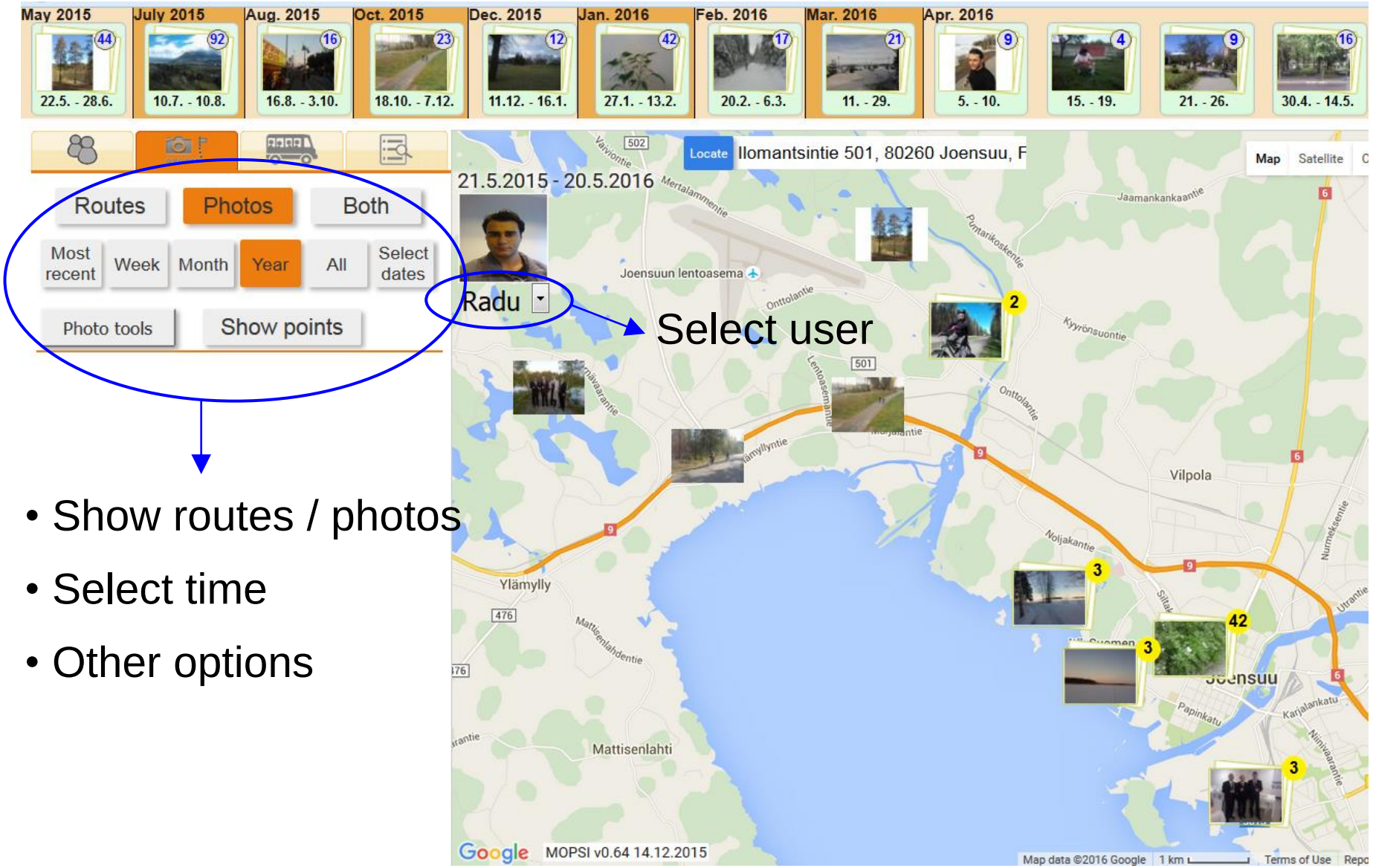






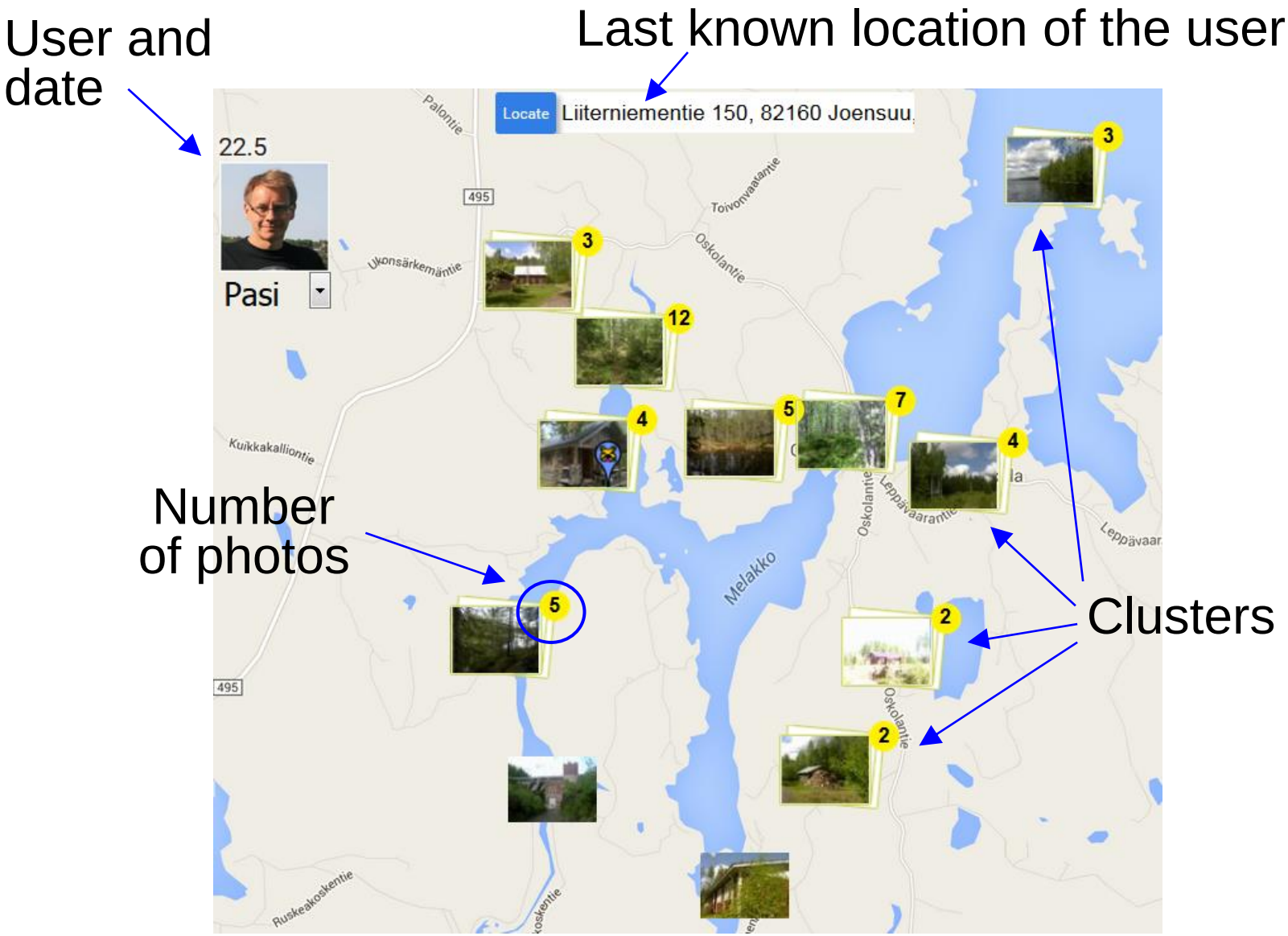
Mopsi website

http://cs.uef.fi/mopsi/

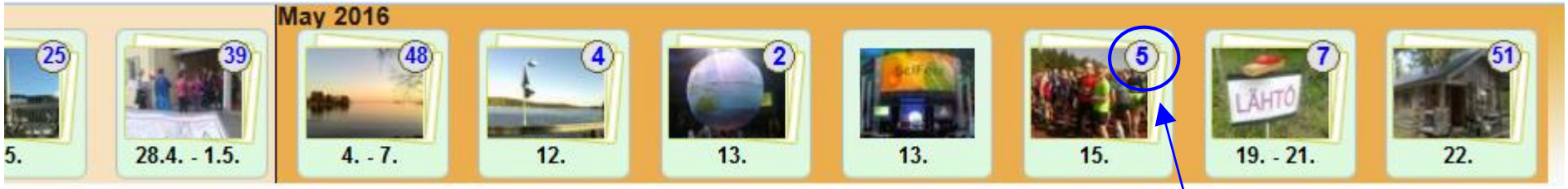


- Show routes / photos
- Select time
- Other options

Photo clusters on map



Clustered timeline view



Clusters

Number of photos

Functions:

Open cluster

Start slideshow

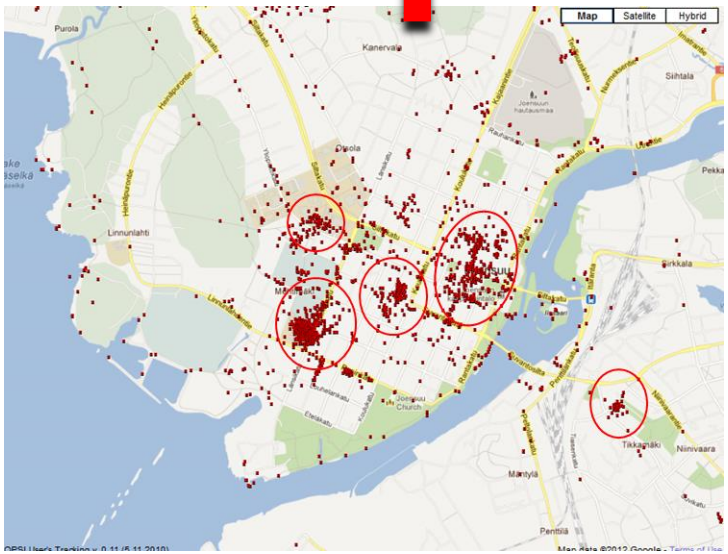
Clustering needed for visualization

Cluttering (too much data)

Timeline view

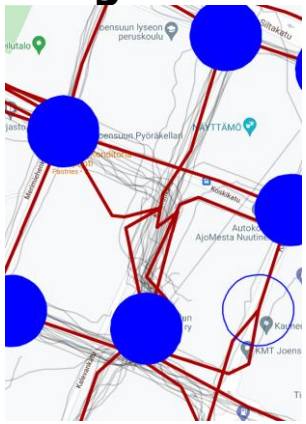
Favorite

Active location



GPS tracks

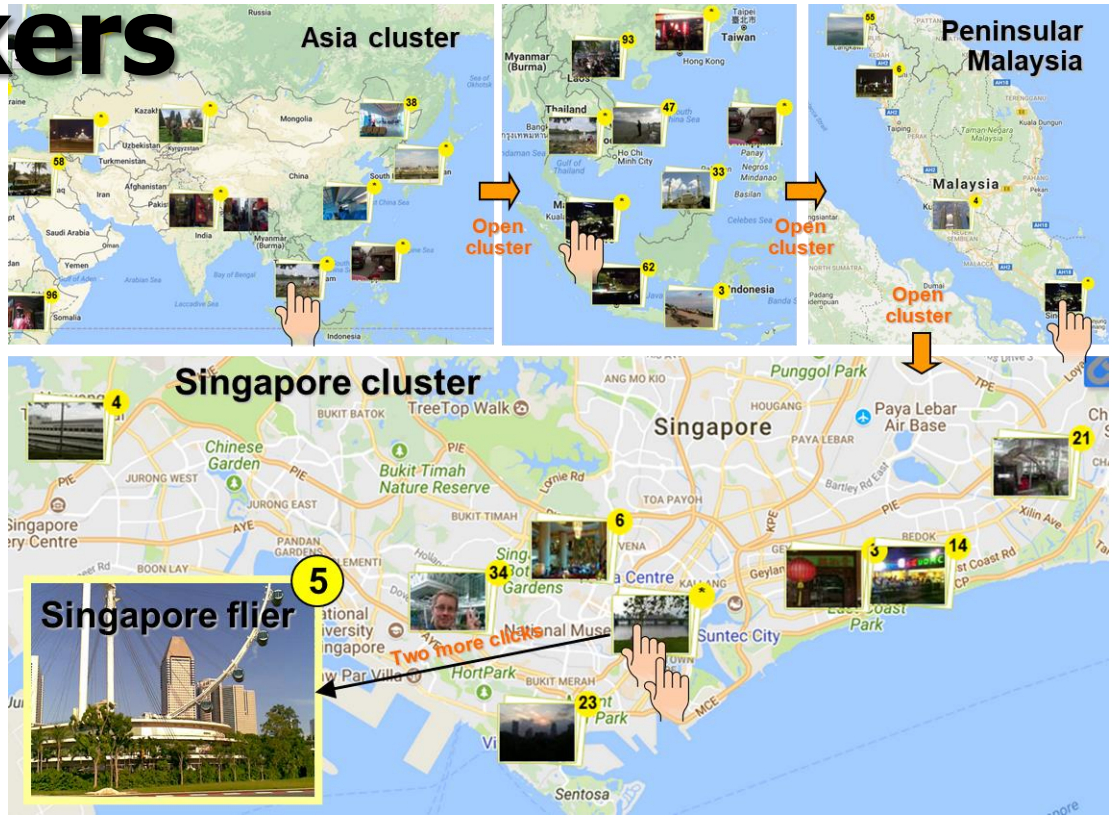
Segments



Clustering markers

Rezaei and Fränti "Real-time clustering of large geo-referenced data for visualizing on map"

Advances in Electrical and Computer Engineering, **2018**



Singapore flier
4 clicks

Tigne Point mall
Malta
3 clicks

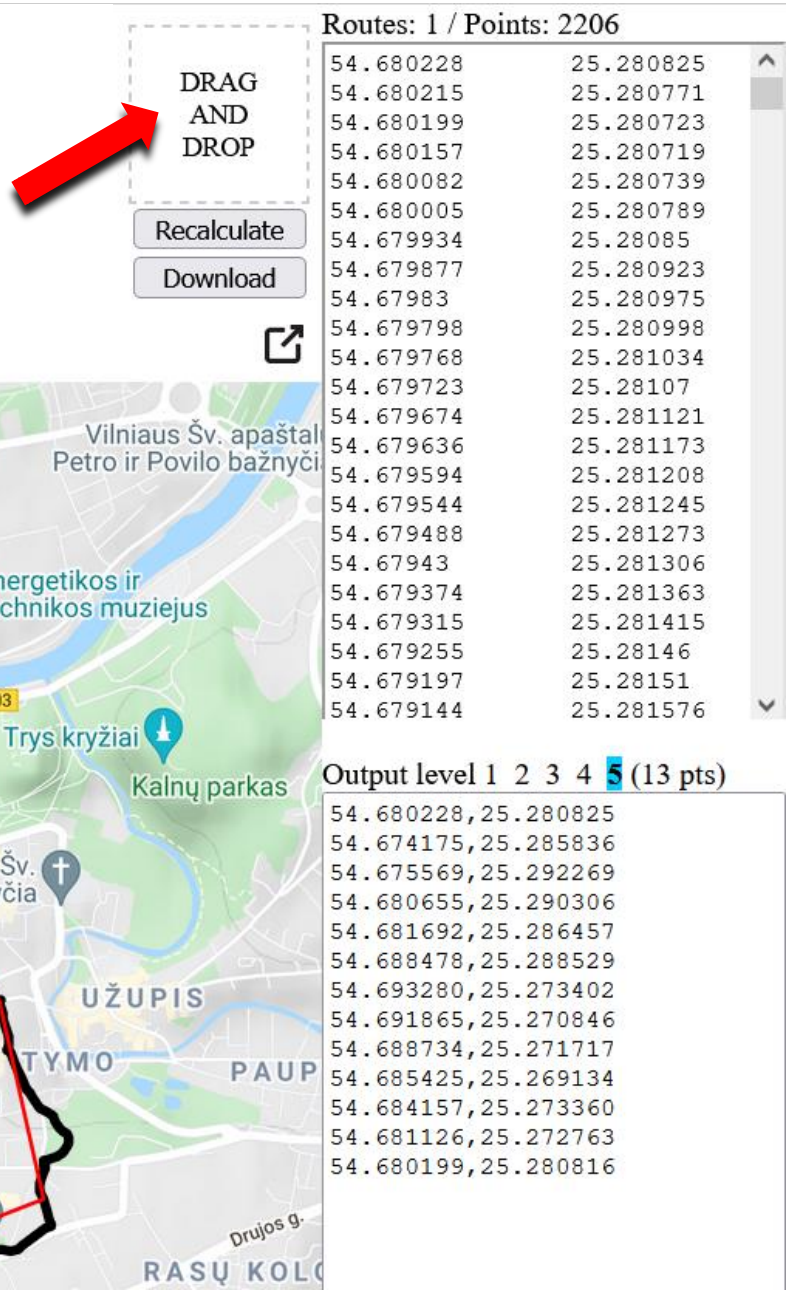
GPS tracks

Route reduction

M. Chen, M. Xu and P. Fränti

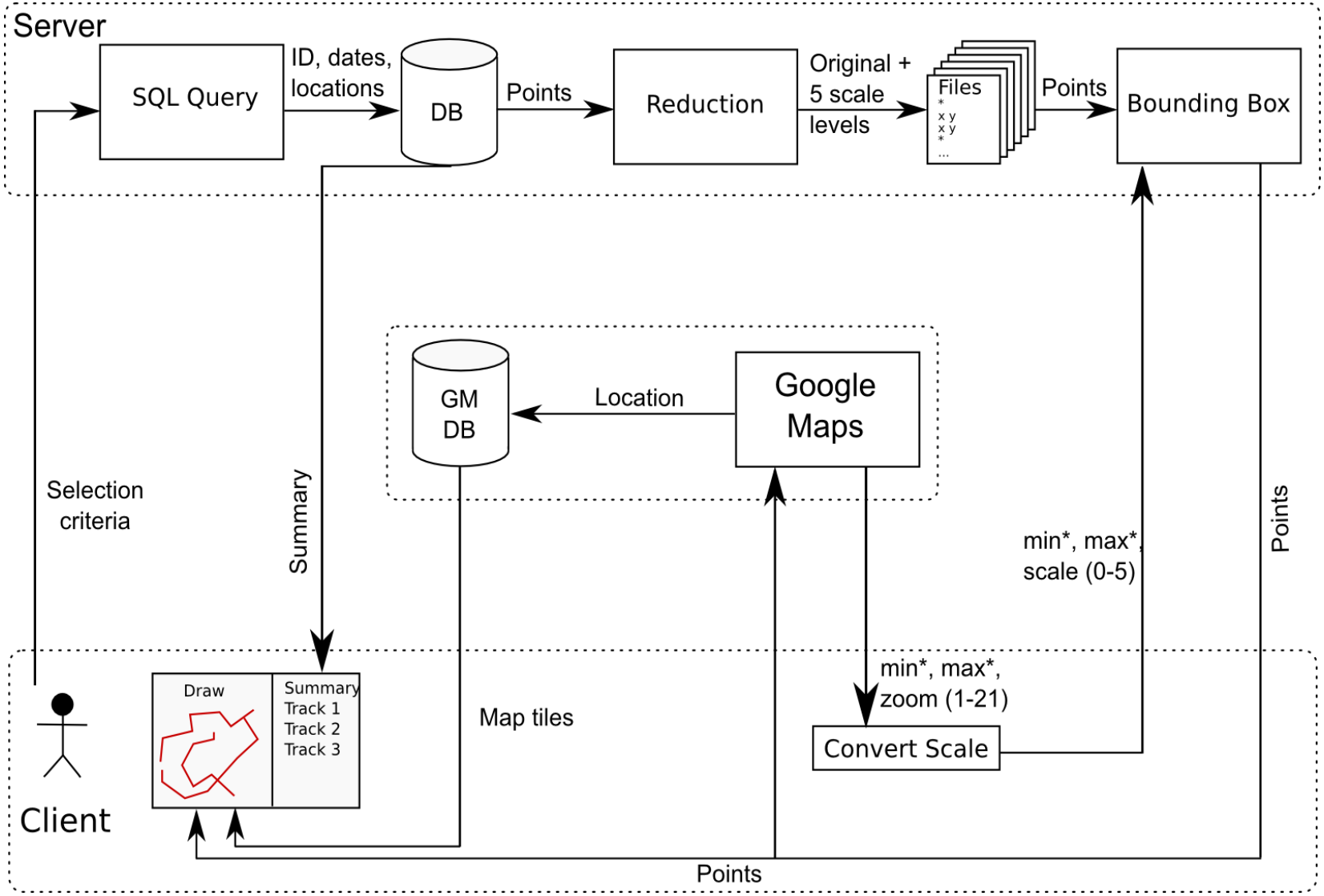
A fast O(N) multi-resolution polygonal approximation algorithm for GPS trajectory simplification

IEEE Trans. on Image Processing, **2012**



Real-time route processing

Waga, Tabarcea, Mariescu-Istodor and Fränti, "Real time access to multiple GPS tracks", *WEBIST,* **2013**



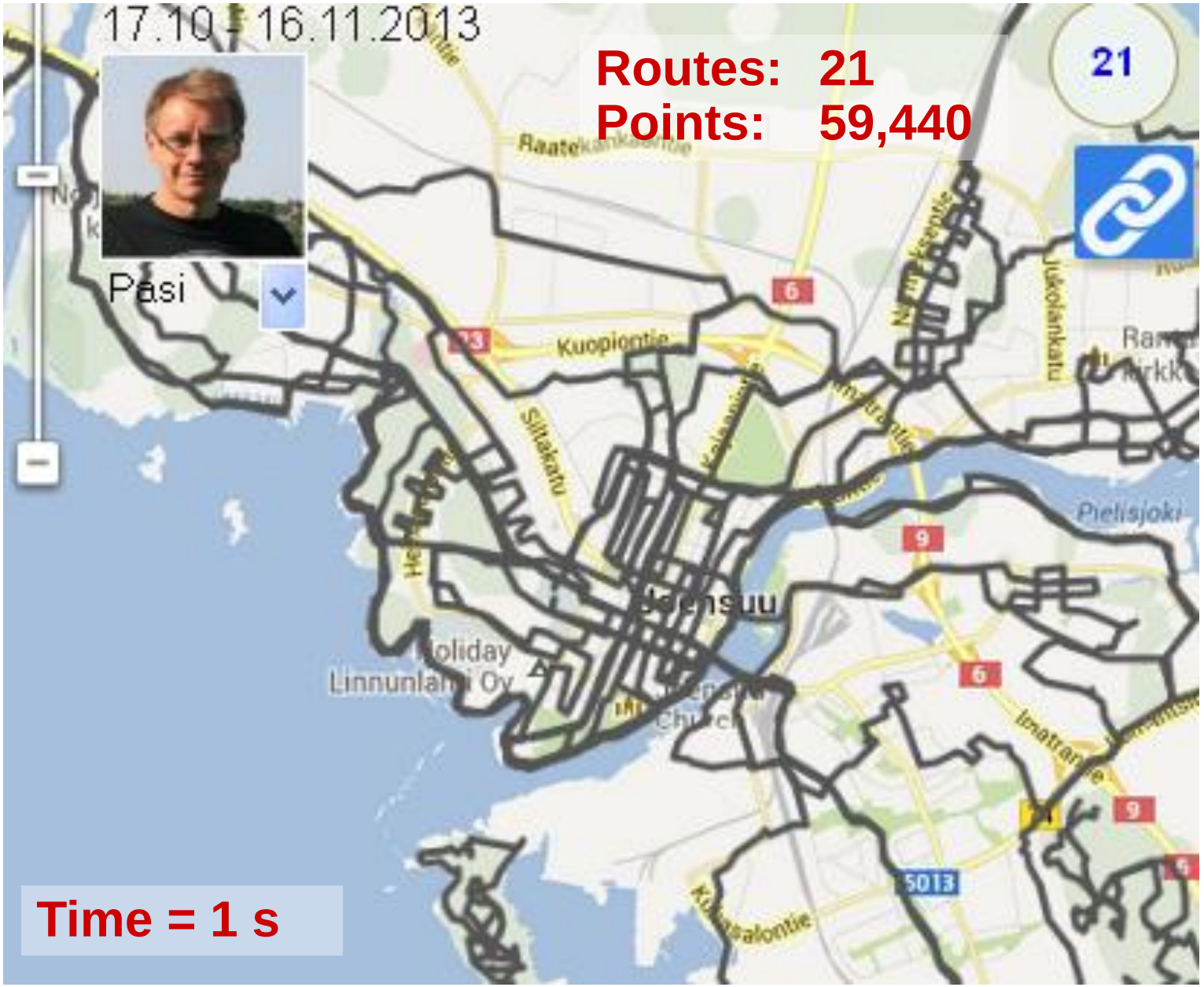
K. Waga

A. Tabarcea

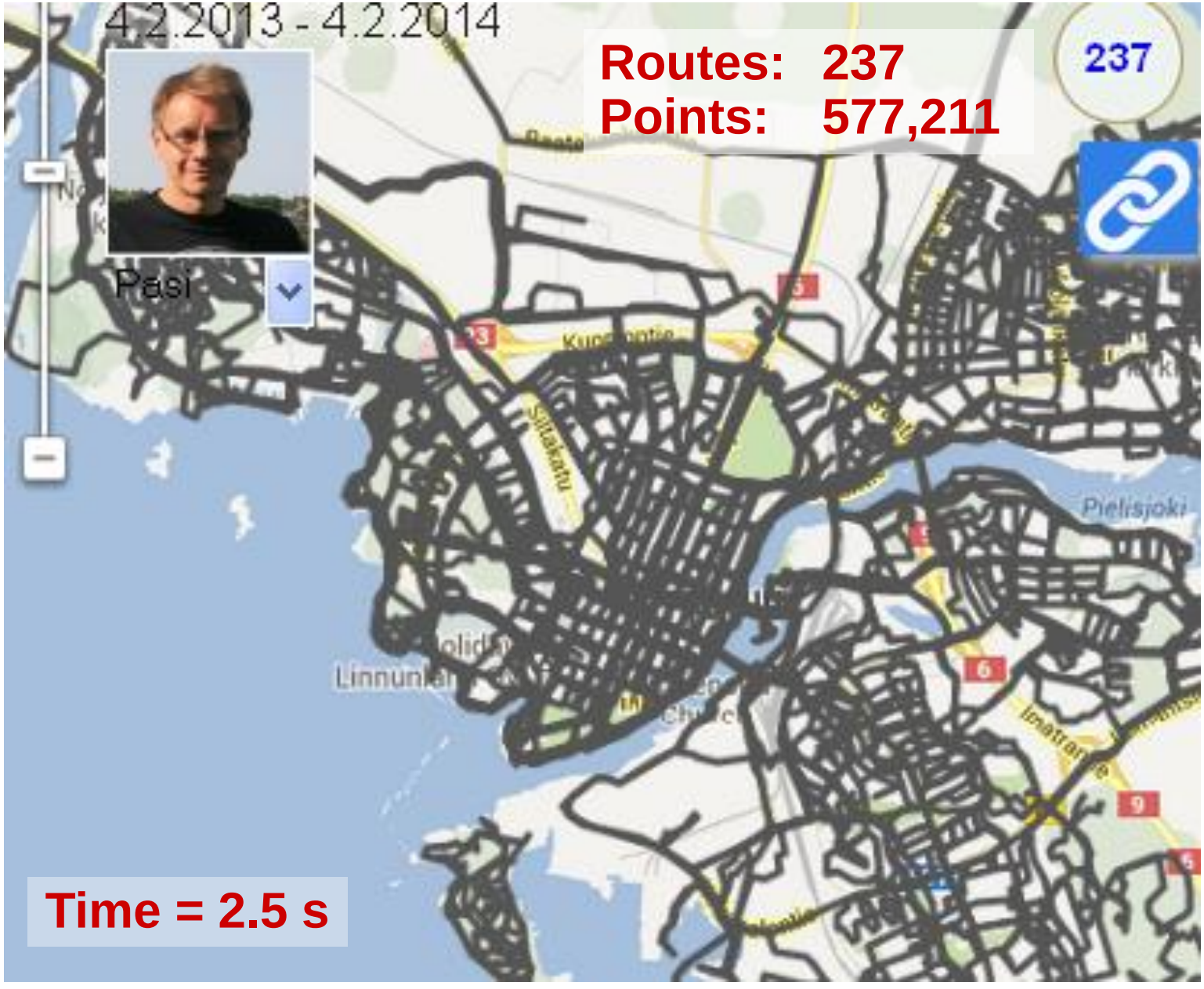
R. Mariescu-Istodor

Routes: one week

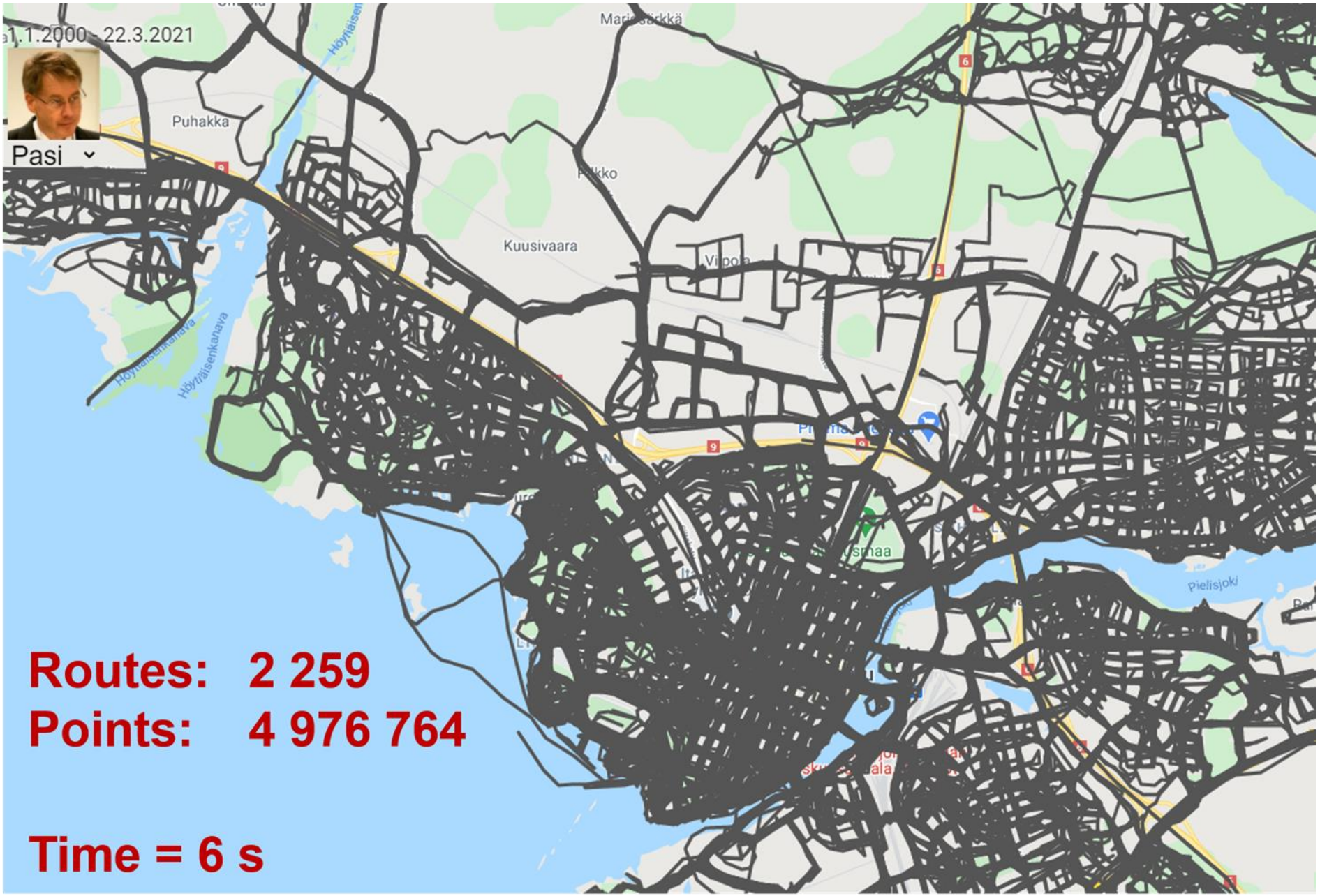
Routes: one month



Routes: one year

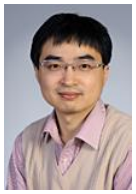


Routes: All time




K. Waga


A. Tabarcea


M. Chen

Move type detection

Waga, Tabarcea, Chen and Fränti,
"Detecting movement type by route segmentation and classification", *CollaborateCom,* 2012.

Move type detection

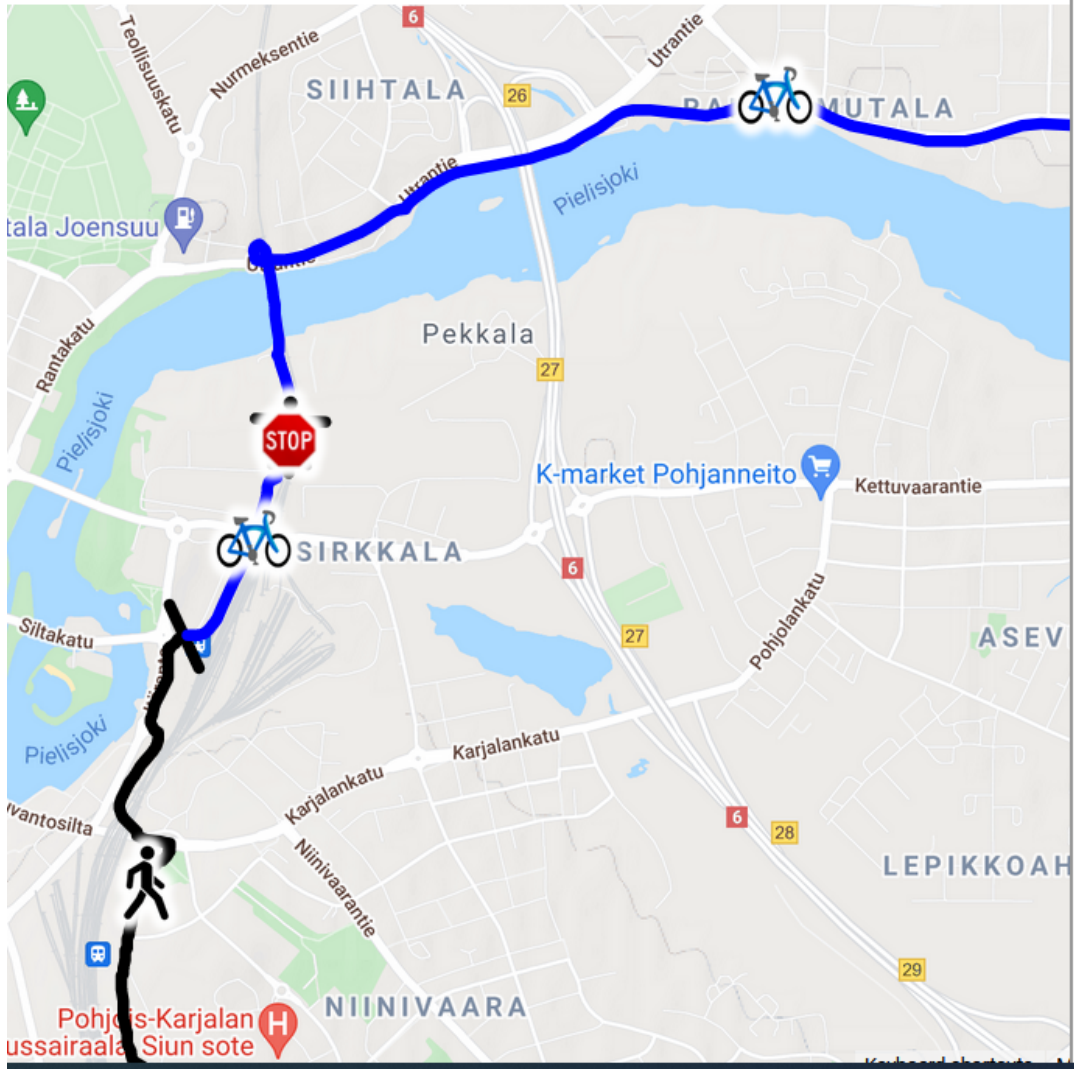
Interactive tool

DRAG AND DROP

Recalculate

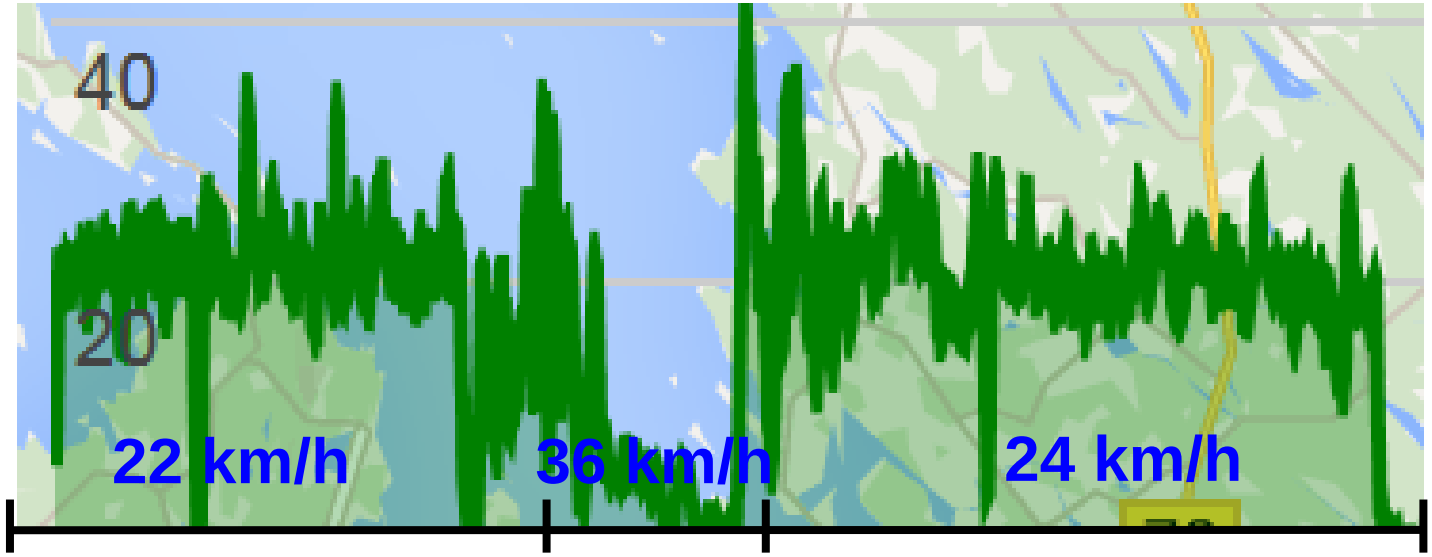
Download

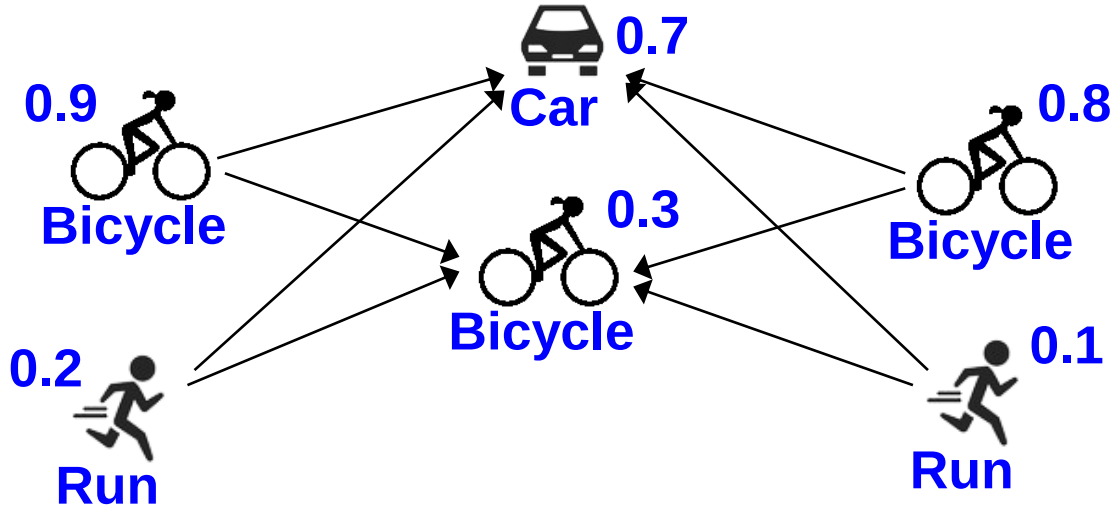
Routes: 1 / Points: 1085




```
62.589065,29.780445,1381444957609
62.589105,29.780401,1381444959620
62.589135,29.780347,1381444961618
62.589167,29.780309,1381444963609
62.589191,29.780275,1381444965617
62.589218,29.780244,1381444967619
62.589243,29.780201,1381444969609
62.589274,29.780162,1381444971616
62.589293,29.780135,1381444973606
62.589309,29.780099,1381444975618
62.589330,29.780058,1381444977616
62.589343,29.780027,1381444979617
62.589342,29.780021,1381444981644
62.589353,29.779997,1381444983619
62.589369,29.779965,1381444985610
62.589386,29.779915,1381444987603
62.589393,29.779860,1381444989598
62.589410,29.779820,1381444991614
62.589434,29.779773,1381444993604
62.589446,29.779722,1381444995604
62.589469,29.779663,1381444997608
62.589476,29.779614,1381444999613
62.589493,29.779571,1381445001613
62.589503,29.779532,1381445003613
62.589522,29.779488,1381445005606
62.589545,29.779451,1381445007603
62.589576,29.779403,1381445009604
62.589607,29.779355,1381445011601
62.589629,29.779313,1381445013605
62.589646,29.779271,1381445015608
62.589659,29.779225,1381445017605
62.589673,29.779178,1381445019618
62.589688,29.779137,1381445021611
62.589706,29.779101,1381445023627
62.589699,29.779071,1381445025617
62.589726,29.779031,1381445027622
62.589741,29.778985,1381445029619
62.589757,29.778910,1381445031615
62.589779,29.778884,1381445033618
62.589793,29.778864,1381445035607
62.589807,29.778827,1381445037611
62.589814,29.778768,1381445039616
62.589830,29.778706,1381445041622
62.589853,29.778657,1381445043618
62.589872,29.778593,1381445045604
62.589885,29.778542,1381445047611
62.589899,29.778493,1381445049604
62.589912,29.778456,1381445051612
62.589929,29.778404,1381445053616
62.589941,29.778352,1381445055613
62.589960,29.778298,1381445057596
62.589962,29.778249,1381445059609
62.589980,29.778206,1381445061613
62.589991,29.778157,1381445063617
62.590009,29.778100,1381445065613
62.590025,29.778052,1381445067608
62.590028,29.777999,1381445069602
62.590043,29.777964,1381445071614
62.590052,29.777891,1381445073602
62.590060,29.777830,1381445075617
```

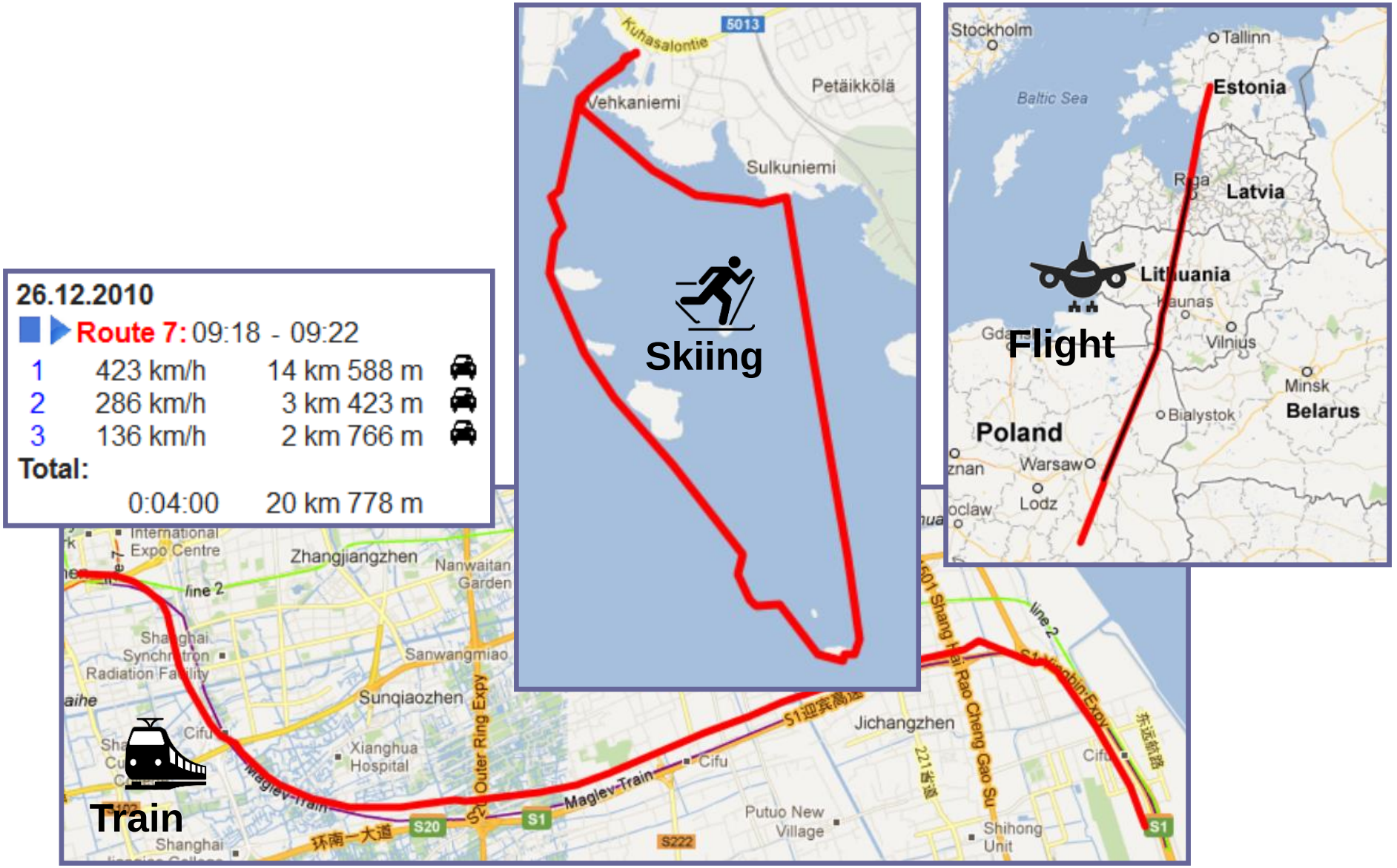
2-direction Hidden Markov Model





Summarization of the activity

Challenging moving types

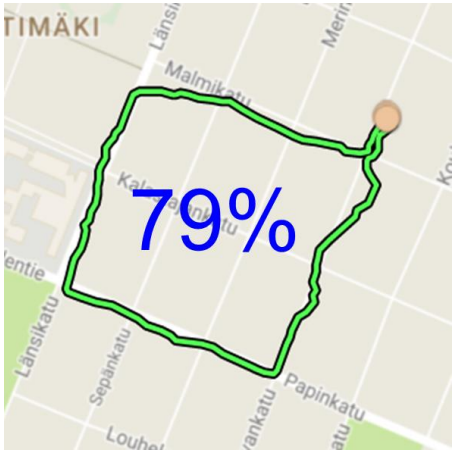


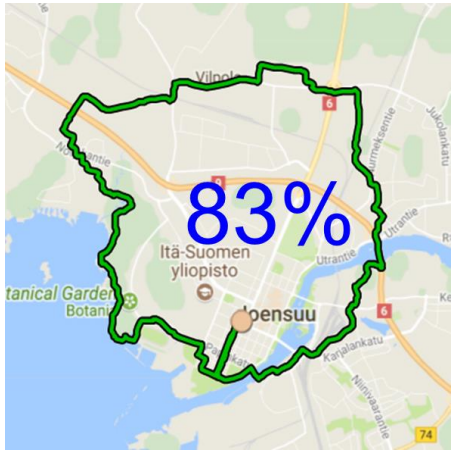
Roundness

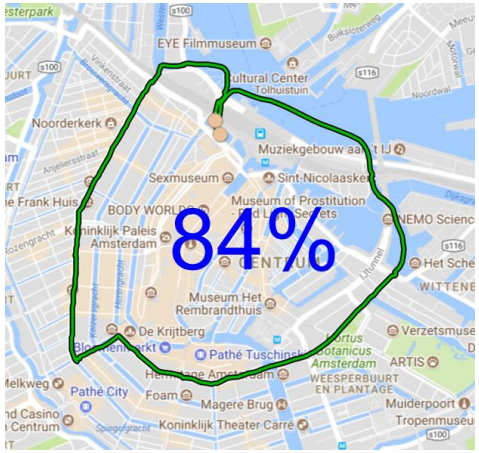

P. Heng

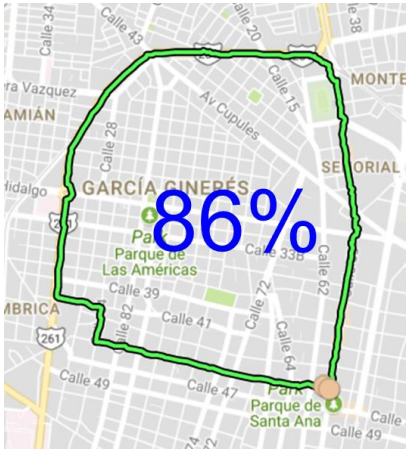

R. Mariescu-Istodor

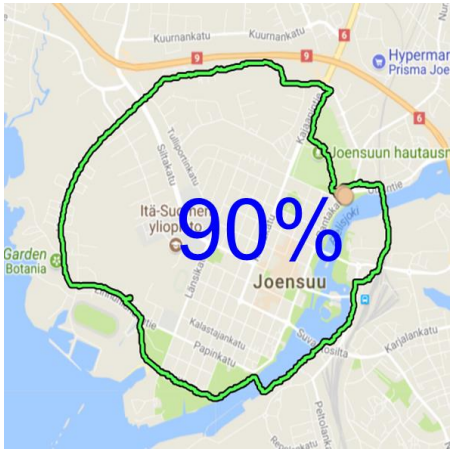
Mariescu-Istodor, Heng and Fränti
Roundness measure for GPS routes
LBS, **2018**

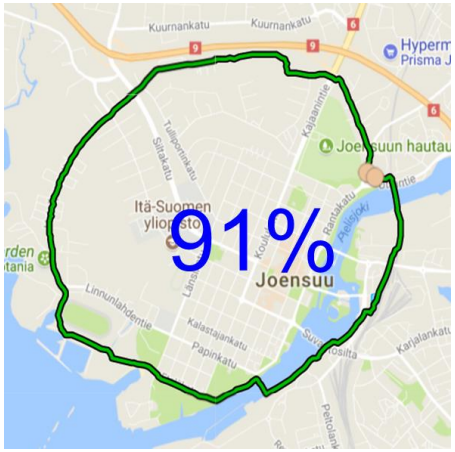






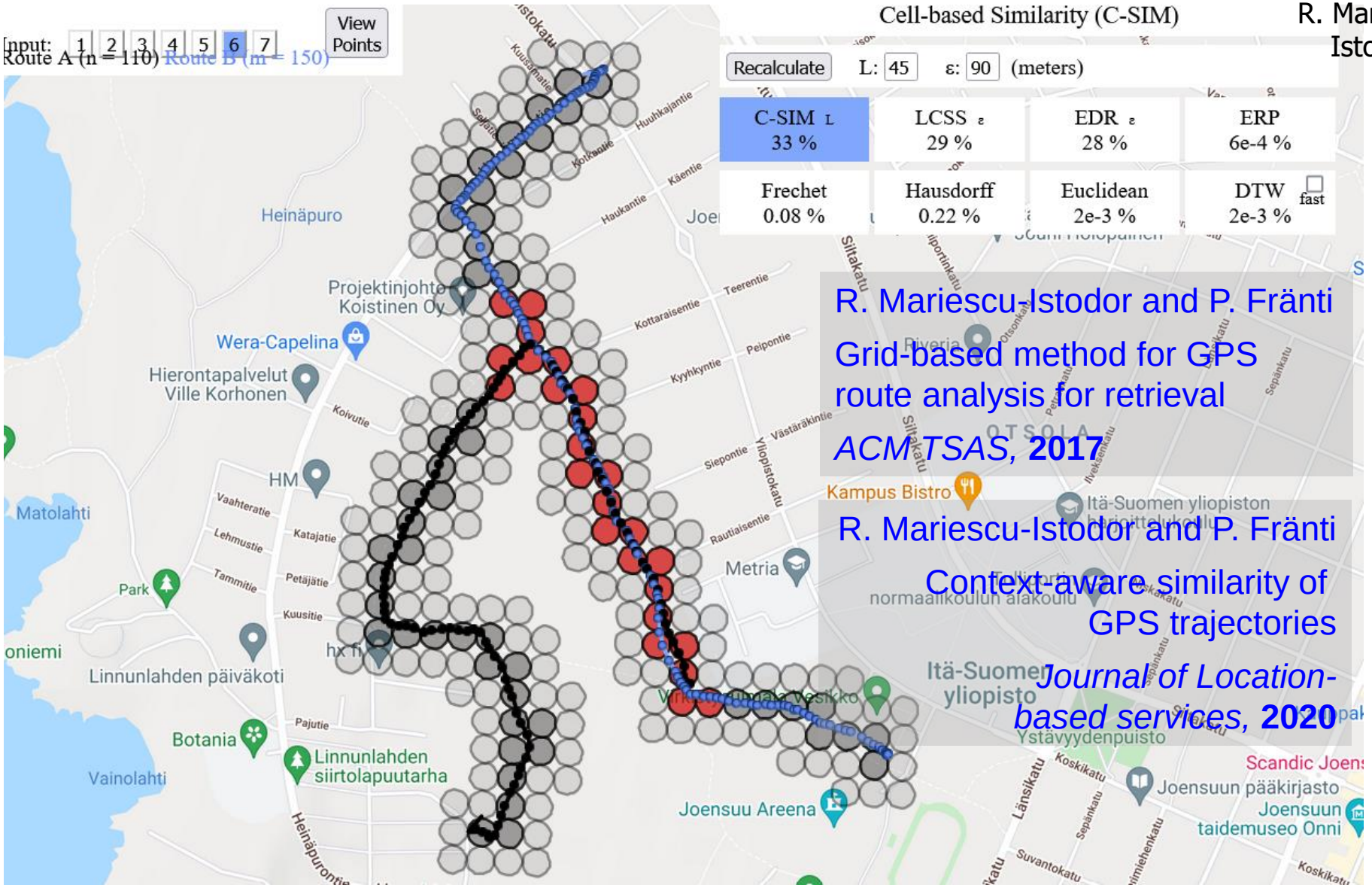






Trajectory similarity

R. Mariescu-Istodor



R. Mariescu-Istodor and P. Fränti

Grid-based method for GPS route analysis for retrieval

ACM TSAS, **2017**

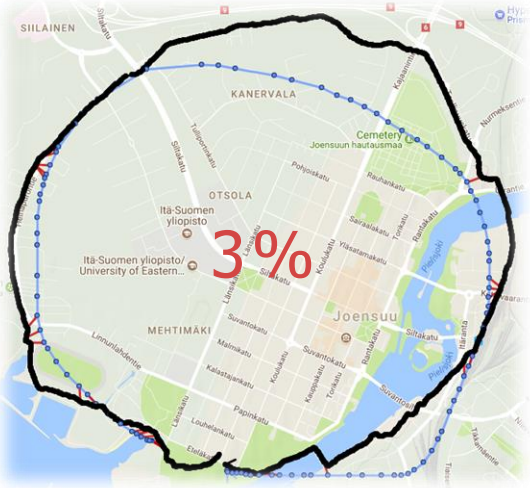
R. Mariescu-Istodor and P. Fränti

Context-aware similarity of GPS trajectories

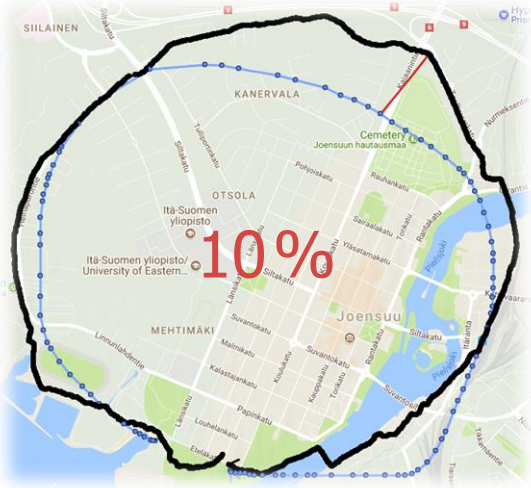
Journal of Location-based services, **2020**

Different Similarity Measures

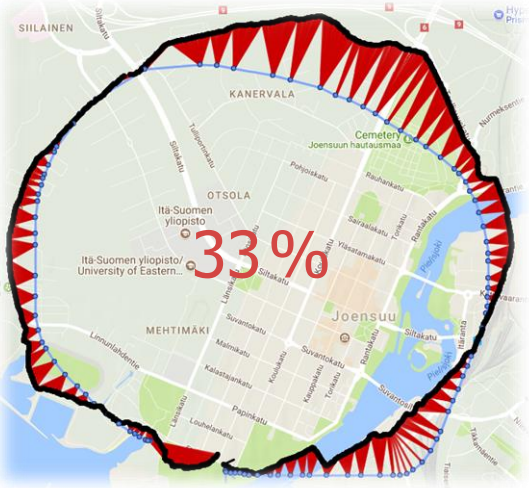
LCSS



Frechet



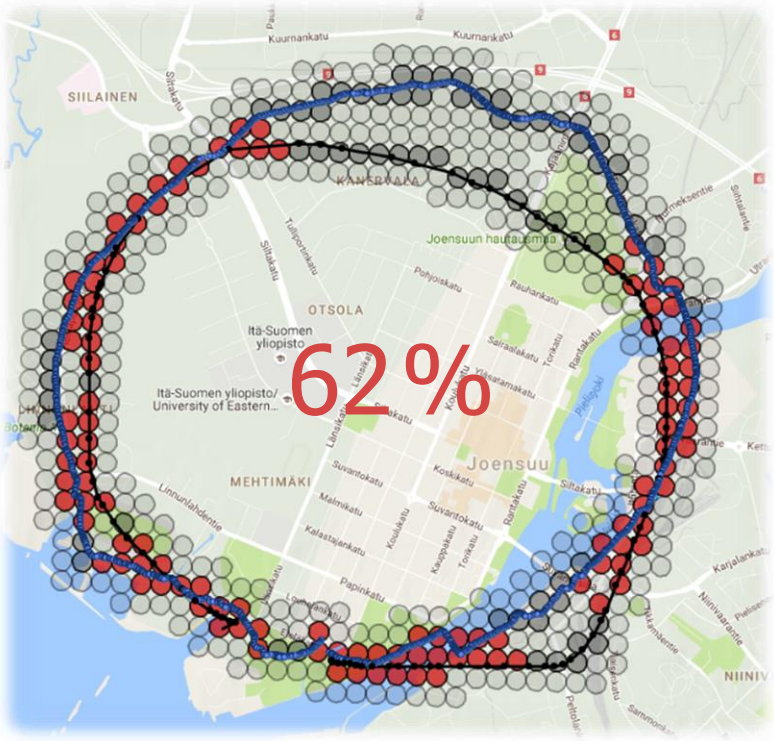
DTW

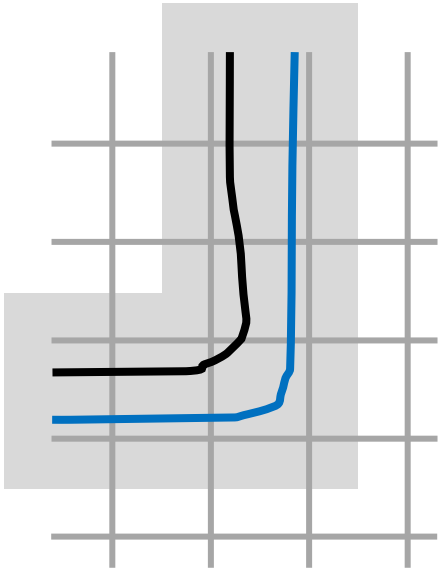


C-SIM

Mariescu-Istodor and Fränti, "Grid-based method for GPS route analysis for retrieval",
ACM Trans. on Spatial Algorithms and Systems, **2017**

R. Mariescu-Istodor





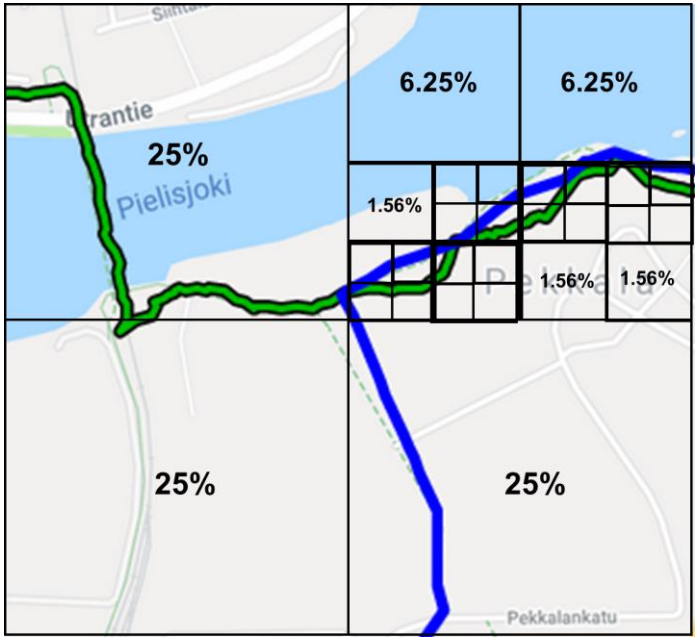
$$S(C_A, C_B) = \frac{|C_A \cap C_B| + |C_A \cap C_B{}^d| + |C_B \cap C_A{}^d|}{|C_A| + |C_B| - |C_A \cap C_B|}$$

R. Mariescu-Istodor

HC-SIM: hierarchical variant

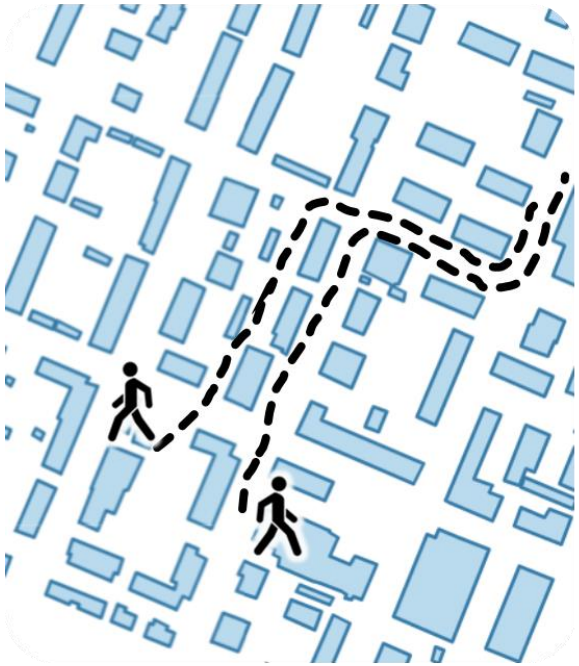
Fränti and Mariescu-Istodor
Averaging GPS segments: competition 2019
Pattern Recognition, **2021**

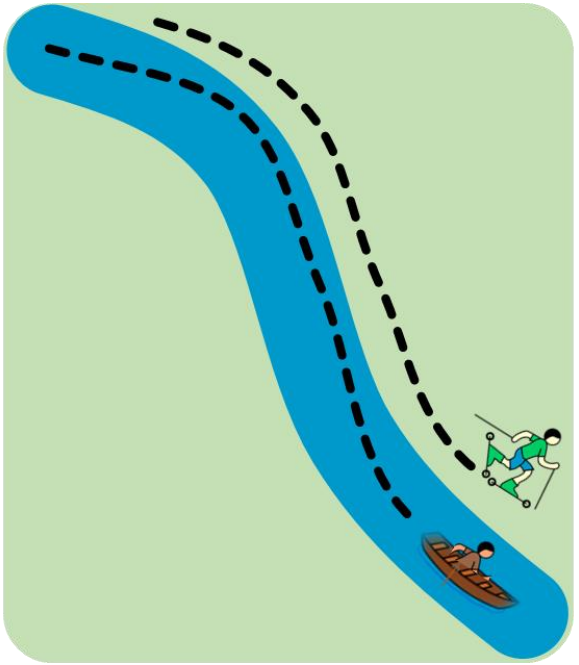
$$\text{HC-SIM}(A,B) = \frac{1}{L}\sum_{i=1}^{L} \text{C-SIM}(A,B,0.005 \times 2^{i-1})$$

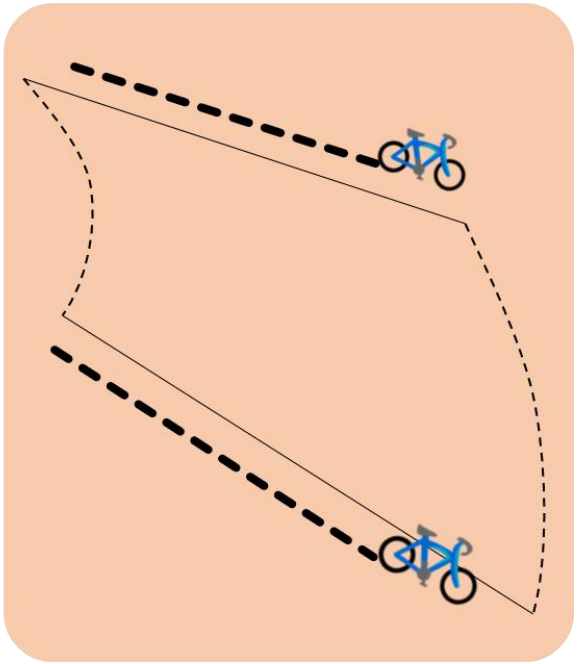


Measure	Correlation
HC-SIM	0.84
C-SIM	0.72
IRD	0.52
LCSS	0.45
EDR	0.37
Hausdorff	0.32
ERP	0.21
DTW	0.11
Euclidean	0.09
Discrete Frechet	0.05

Context-aware similarity







R. Mariescu-Istodor

Radu Mariescu-Istodor and Pasi Fränti
Context-aware similarity of GPS trajectories
Journal of Location Based Services, **2020**

Gesture search

R. Mariescu-Istodor

Mariescu-Istodor and Fränti
Gesture input for GPS route search
S+SSPR, **2016**

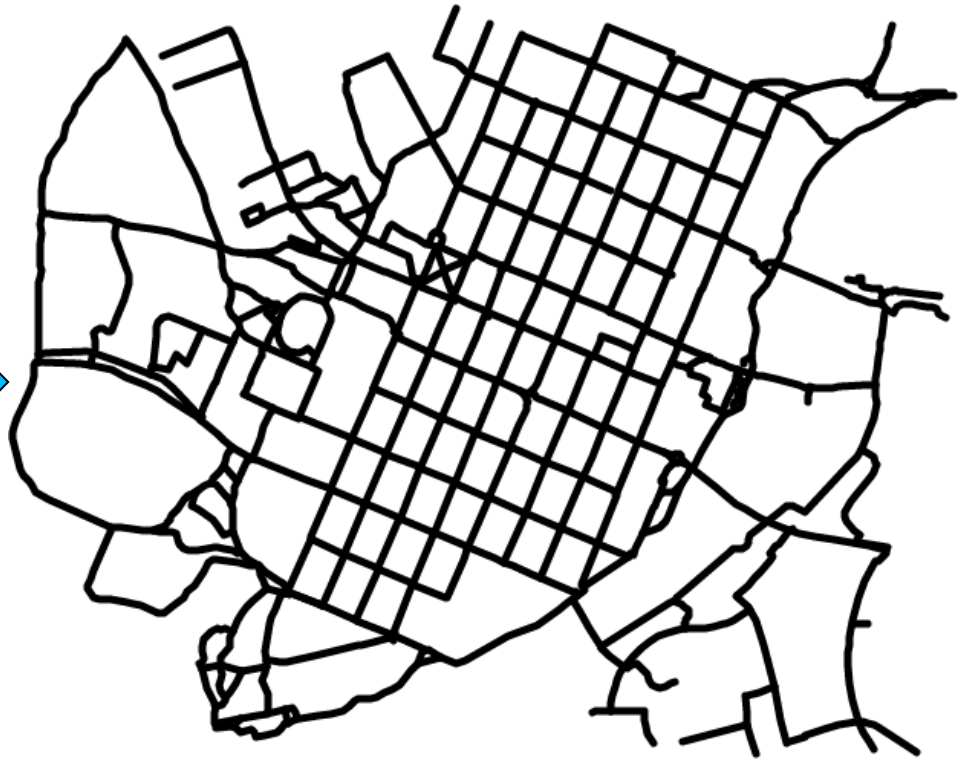
Extracting road network

R. Mariescu-Istodor

R. Mariescu-Istodor and P. Fränti
CellNet: Inferring road networks from GPS trajectories,
ACM TSAS, **2018**

GPS trajectories

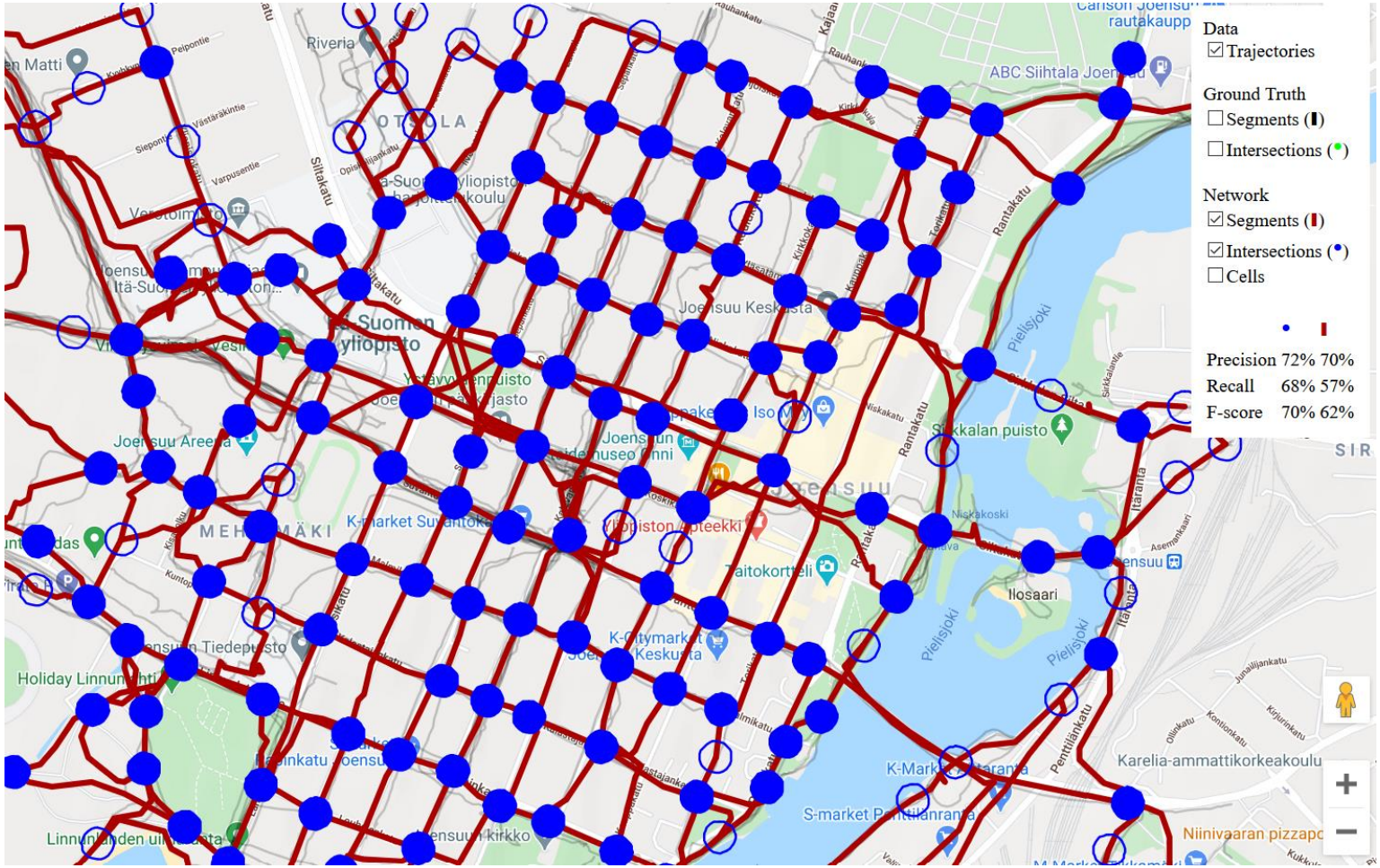
Road network



Interactive tool + data

Joensuu & Chicago

http://cs.uef.fi/sipu/mopsi/
http://cs.uef.fi/mopsi/routes/network/



Interactive tool + data

1. CellNet

http://cs.uef.fi/sipu/mopsi/
http://cs.uef.fi/mopsi/routes/network/

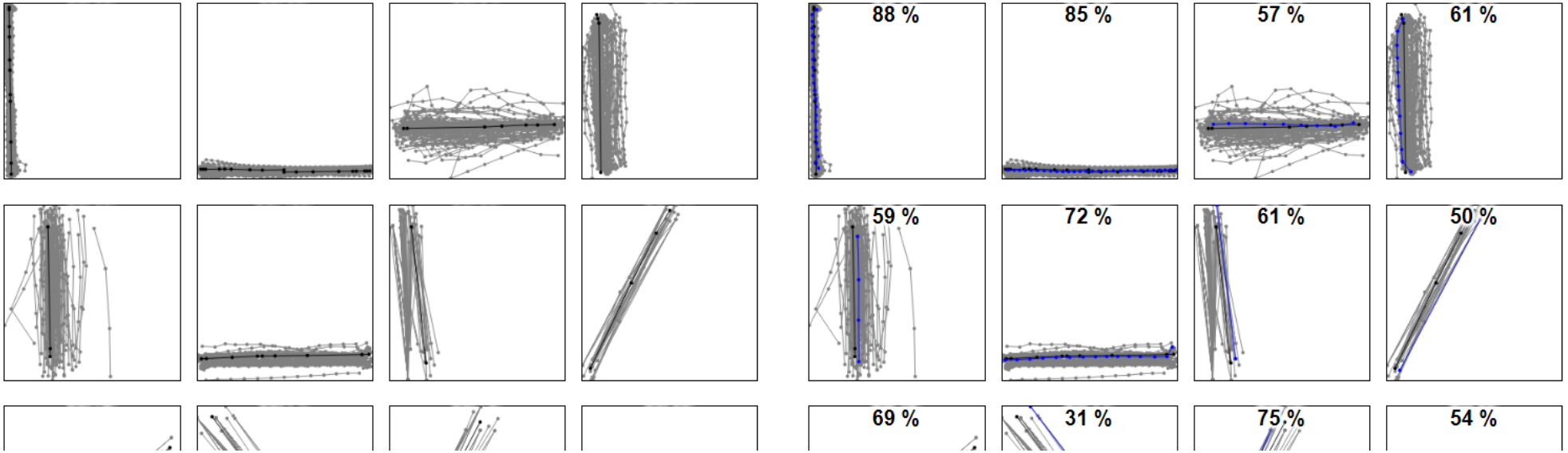
2. Jonathan Davies, Alastair R. Beresford and Andy Hopper. 2006. Scalable, distributed, real-time map generation. *IEEE Pervasive Computing*, 5(4), pp. 47-54.

3. Stefan Edelkamp and Stefan Schrödl. 2003. Route planning and map inference with global positioning traces. *In Computer Science in Perspective*, pp. 128-151.

Segments averaging competition

Pasi Fränti and Radu Mariescu-Istodor, "Averaging GPS segments competition 2019",
Pattern Recognition, **2020**

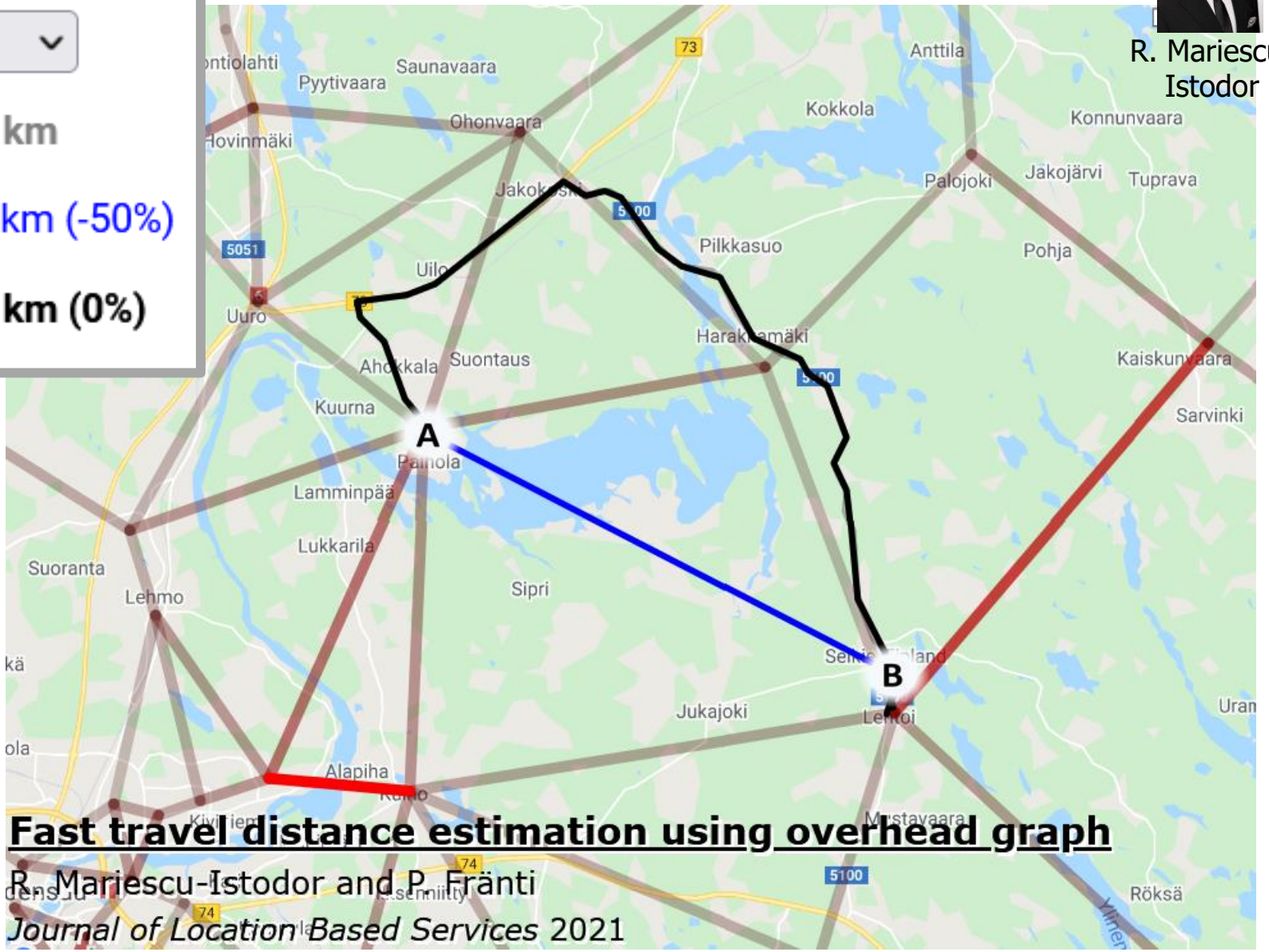
R. Mariescu-Istodor



Distance estimation

R. Mariescu-Istodor

Nodes	256
True:	23.2 km
Bird:	11.5 km (-50%)
Estimated: 23.2 km (0%)	



R. Mariescu-Istodor
P. Fränti

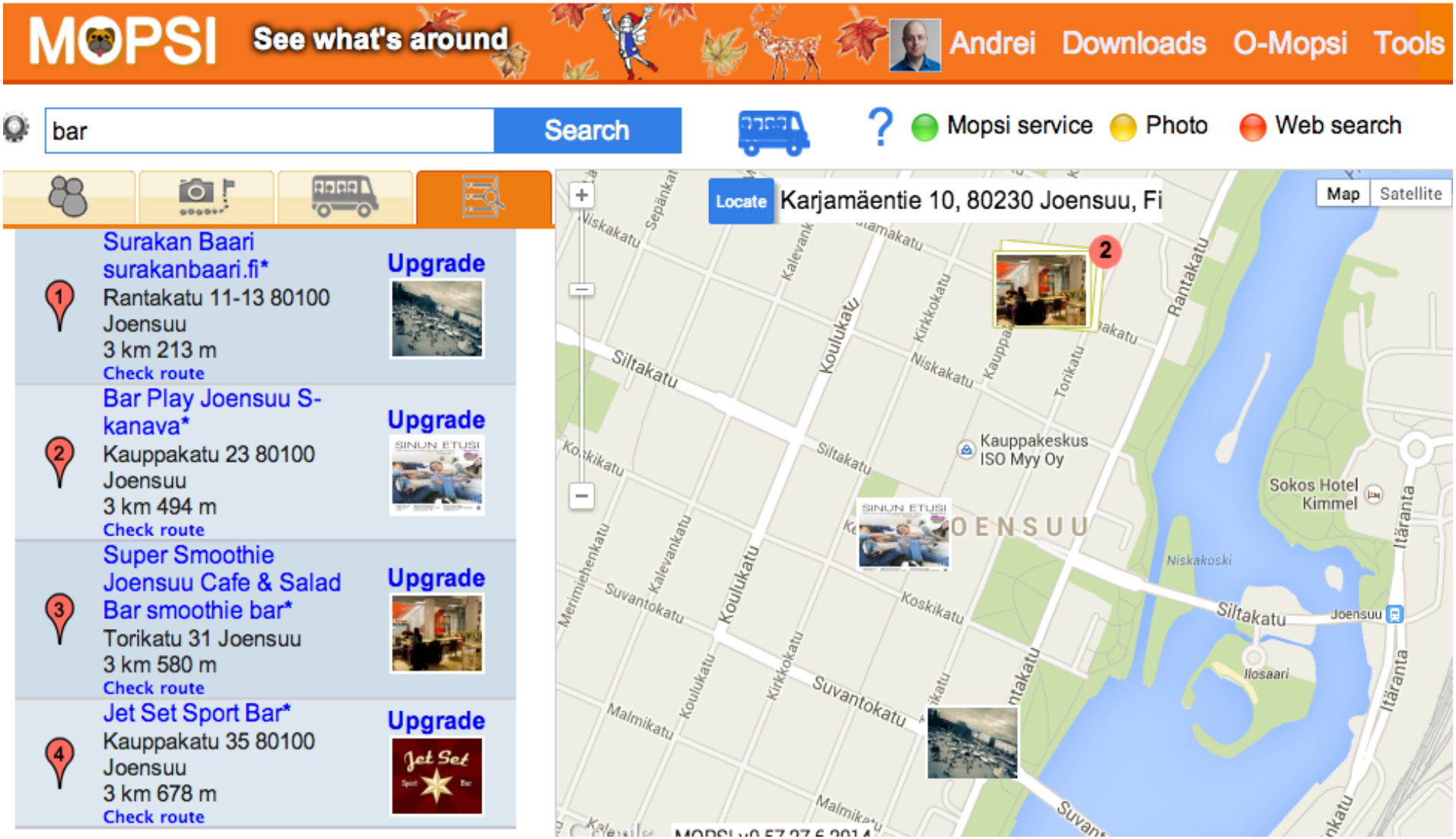
Fast travel distance estimation using overhead graph

Journal of Location-Based Services
2021

Mopsi search

Mopsi search

Tabarcea, Gali and Fränti, "Framework for location-aware search engine", *Journal of Location Based Services*, **2017**



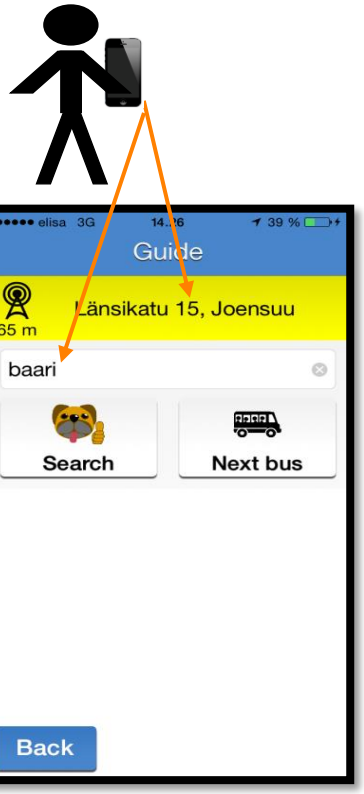
General workflow

meta search engine

N. Gali

A. Tabarcea

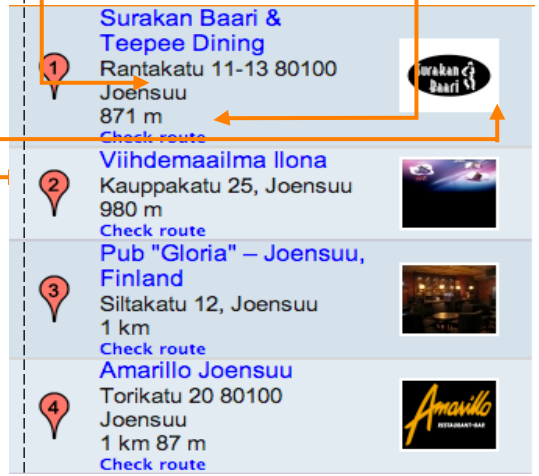
User input



Web mining

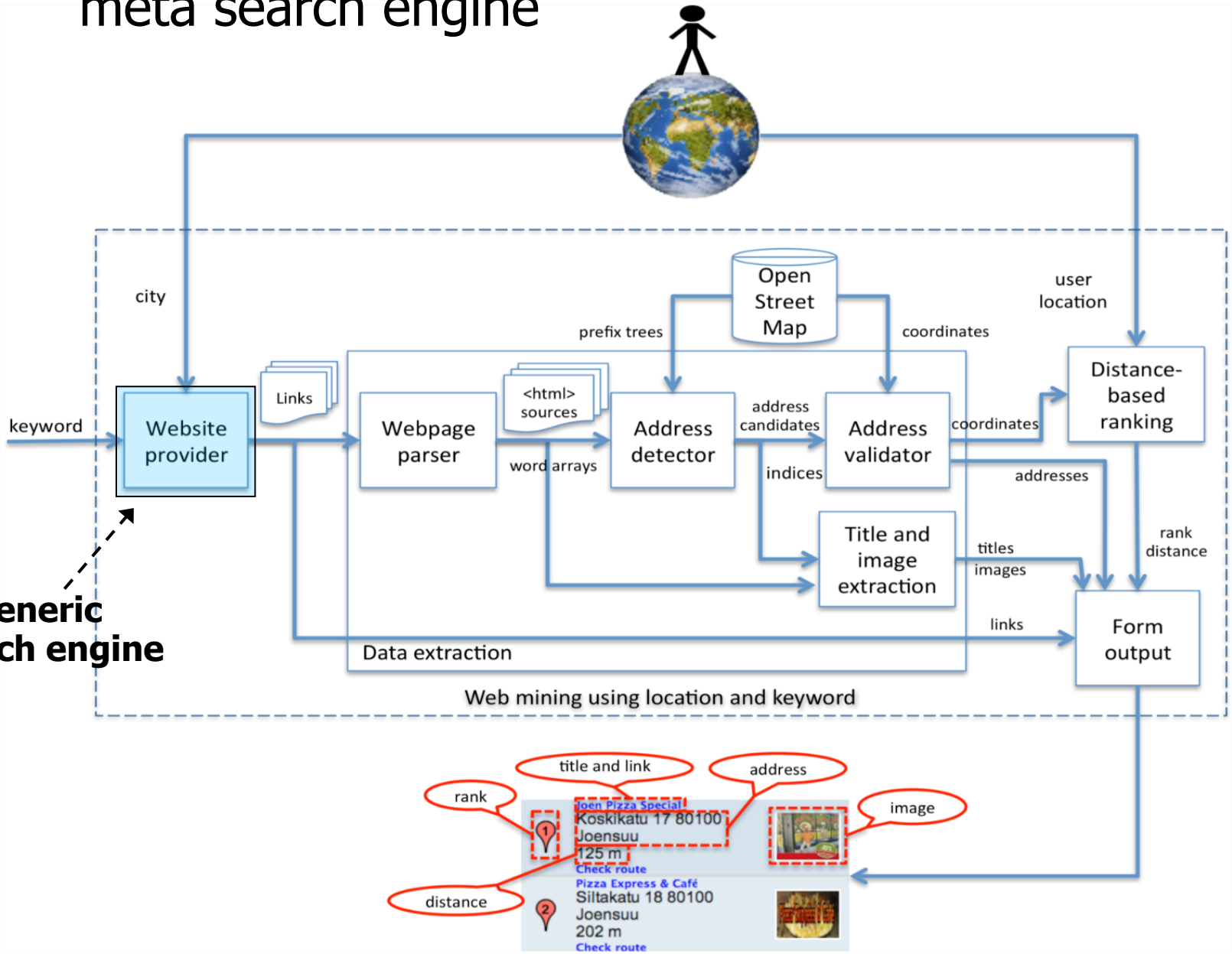
Formatted output

Distance from user



System architecture

meta search engine



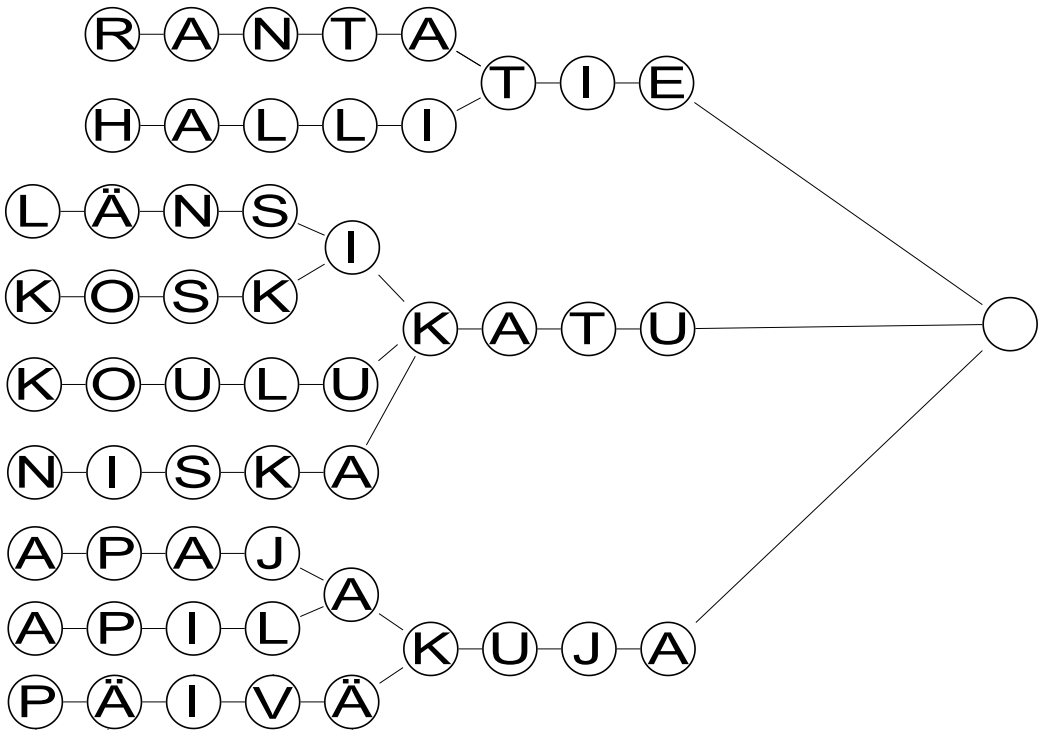
Detecting address from web

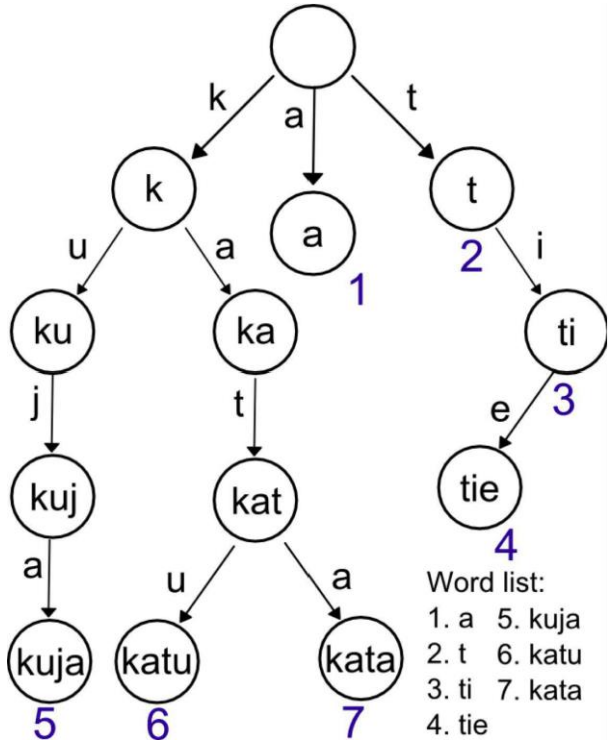
Tabarcea, Hautamäki, Fränti,
Ad-hoc georeferencing of web-pages using street-name prefix trees
WEBIST **2010**


A. Tabarcea


V. Hautamäki

- Analysis of text content of web page
- Matching strings with address databased
- Address database stored as **prefix tree**
- Both street number and postal code required

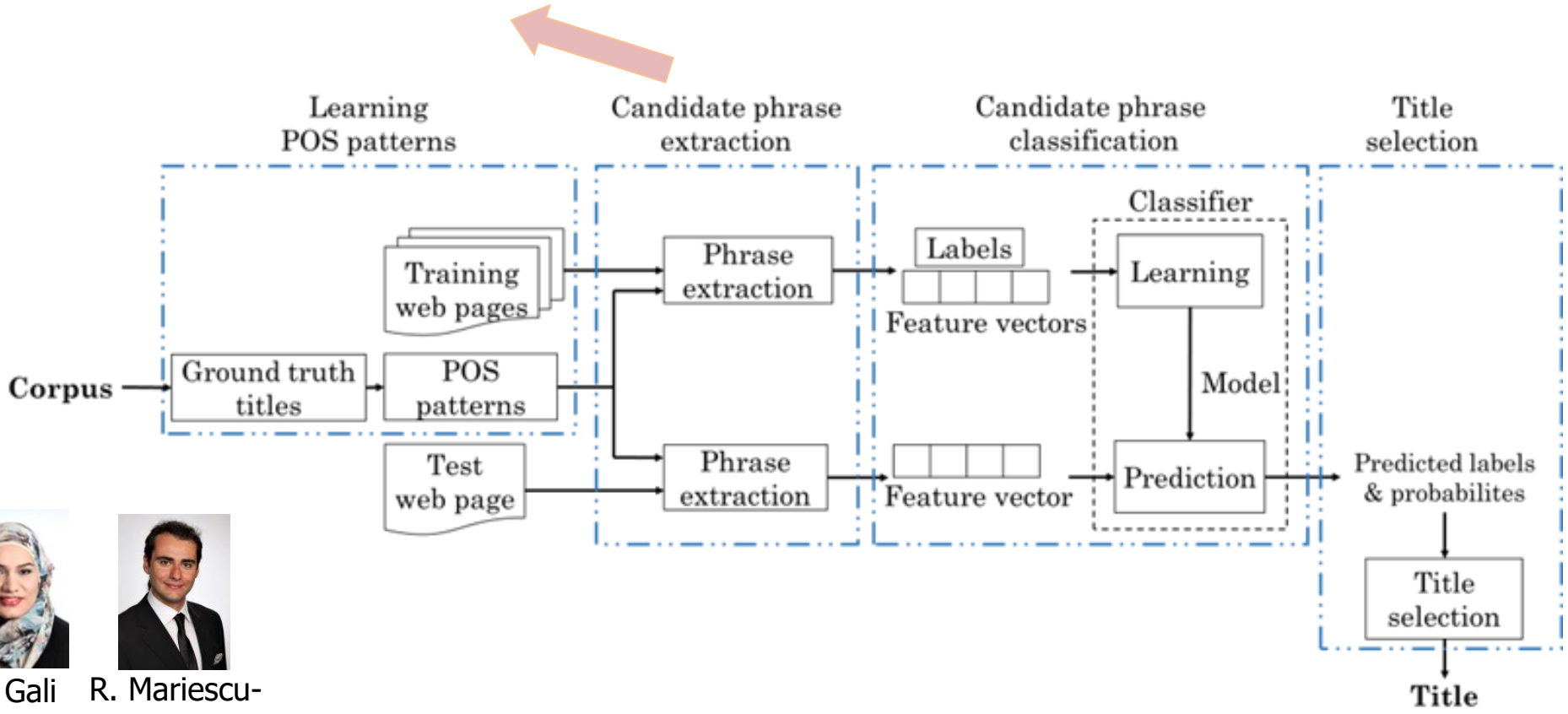




Representative title

Gali, Mariescu-Istodor and Fränti, "Using linguistic features to automatically extract web page title", *Expert Systems with Applications*, **2017**.

Content of text nodes ⇨ N-grams (n=1…6) ⇨ Filter by part-of-speech (POS) patterns



N. Gali

R. Mariescu-Istodor

Keywords


M. Rezaei N. Gali

Web page

Text extraction and pre-processing

POS tagging and nouns extraction

Nouns lemmatization

Nouns Similarities

Nouns clustering and clusters ranking

Keywords selection

Part of the extracted text

<h1>ABOUT FORME SPA</h1>
Forme Spa offers a tranquil environment designed for relaxation and rejuvenation. The spa is located in ….

POS tagging

Forme/NNP Spa/NNP offers/VBZ a/DT tranquil/JJ environment/NN designed/VBN for/IN relaxation/NN and/CC

Extracted lemmas

Forme, spa, building, treatment, massage, therapy

Similarity matrix of lemmas

	Spa	Building	Treatment	Massage	Therapy
Spa	1.00	0.89	0.23	0.20	0.19
Building	0.89	1.00	0.70	0.67	0.63
Treatment	0.23	0.70	1.00	0.95	0.91
Massage	0.20	0.67	0.95	1.00	0.87
Therapy	0.19	0.63	0.91	0.87	1.00

Complete-link clustering and ranked clusters

11	**Cluster 1:**	**Spa** (33)	Building (1)	
8	**Cluster 2:**	**Treatment** (20)	**Massage** (7)	Therapy (2)
5	**Cluster 3:**	**Auckland** (7)	**Wellington** (6)	City (2)
2	Cluster 4:	Service (5)	Care (2)	

Keywords

Spa, Treatment, Massage, Auckland, Wellington

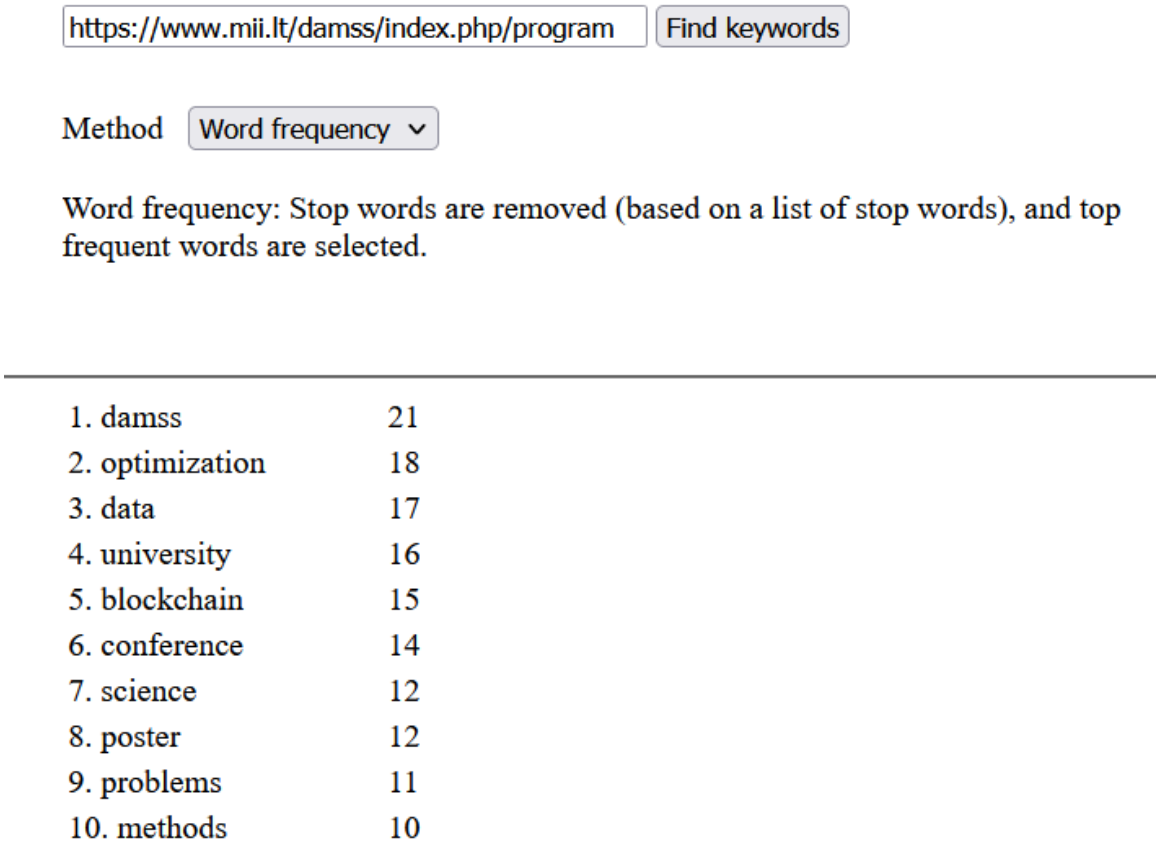
Keyword extraction

M. Rezaei, N. Gali and P. Fränti, ClRank: a method for keyword extraction from web pages using clustering and distribution of nouns, *WI-AIT'15,* **2015**

H. Shah, R. Mariescu-Istodor and P. Fränti, WebRank: Language-independent extraction of keywords from webpages, *IEEE PIC-2021,* **2021**

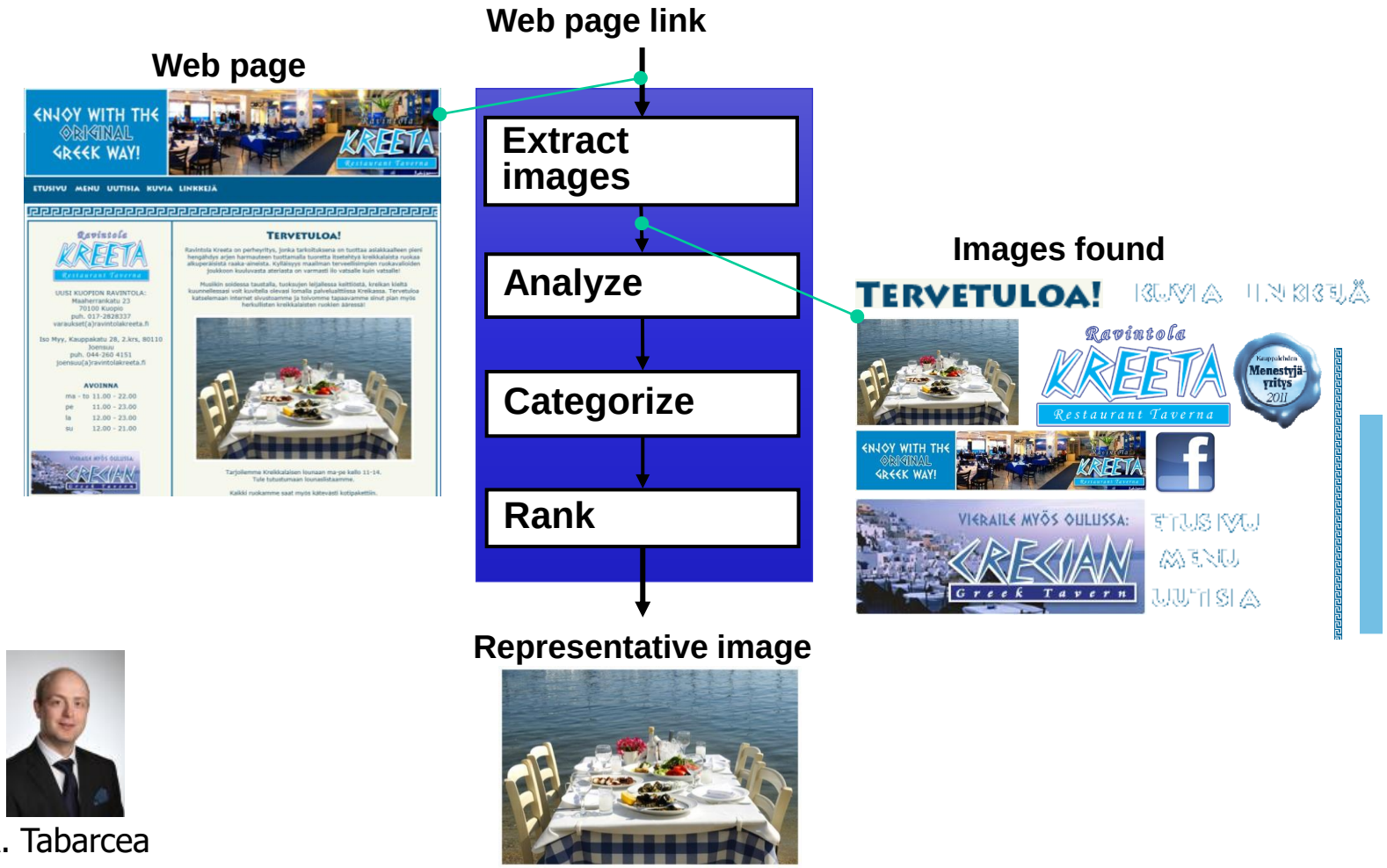
H. Shah, M. Rezaei and P. Fränti, "DOM-based keyword extraction from web pages", *ACM AIIPCC,* **2019**

H. Shah, M.U.S. Khan and P. Fränti, "H-rank: a keywords extraction method from web pages using POS tags", *IEEE INDIN*, **2019**



Representative image

Gali, Tabarcea, and Fränti, "Extracting representative image from web page", *Int. Conf. on Web Information Systems & Technologies*, May **2015**

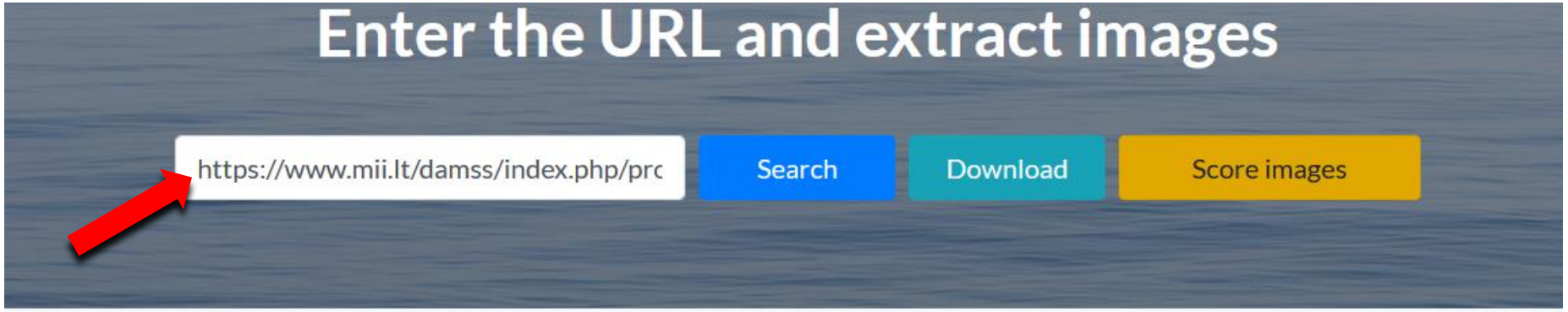



N. Gali

A. Tabarcea

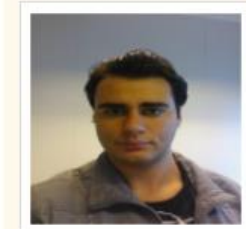
Photo extraction tool

N. Gali, A. Tabarcea, and P. Fränti,
Extracting representative image from web page“
WEBIST **2015**



1		3	619	663	0.93	html	jpg	Representative
2		3	346	346	1.00	html	jpg	Representative

Activity summary



18 h 55 mins

At: Pohjois-Karjala, Penttilä, Joensuu, Finland
Status: Mopsi Maniac
User since: 30.11.2011
Last login: 19.5.2016

1848 Photos
1202 Routes
Favorite

Photo | Routes | Meeting | Check In (7/123)

Annually 14.5.2016

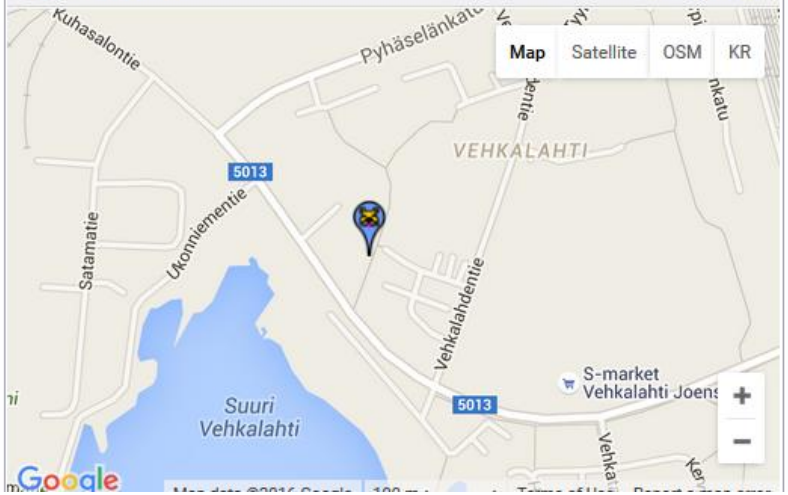
Sipu Cycli...	Nurmes inf...	Waterfall	Rocks	Still snow	Frozen lak...	Skiing slo...
Utra	Finland	Finland	Finland	Finland	Finland	Finland







Recent Data Maximize



Recent Activity Detail

15.5.
11:57 - Radu and matti met
12:10 - Passed by Science Park
12:17 - Radu and matti met
14:08 - Passed by KELA
14:11 - Passed by Timanttiset
14:14 - Passed by Ravintola Rosso
14:25 - Passed by WorldFlicks

14.5.
15:28 - Radu and matti met
17:08 - Took a photo, Sipu Cycling ++

Direct actions

Mariescu-Istodor and Fränti, "Detecting user actions in location-based systems", *LBS* **2018**

R. Mariescu-Istodor

Visiting restaurant


Radu visited Lounasravintola Kuutti.


Radu completed cycling 40 Km.


Radu took a photo *Evening Sky.*


Radu and Pasi met.

Meeting

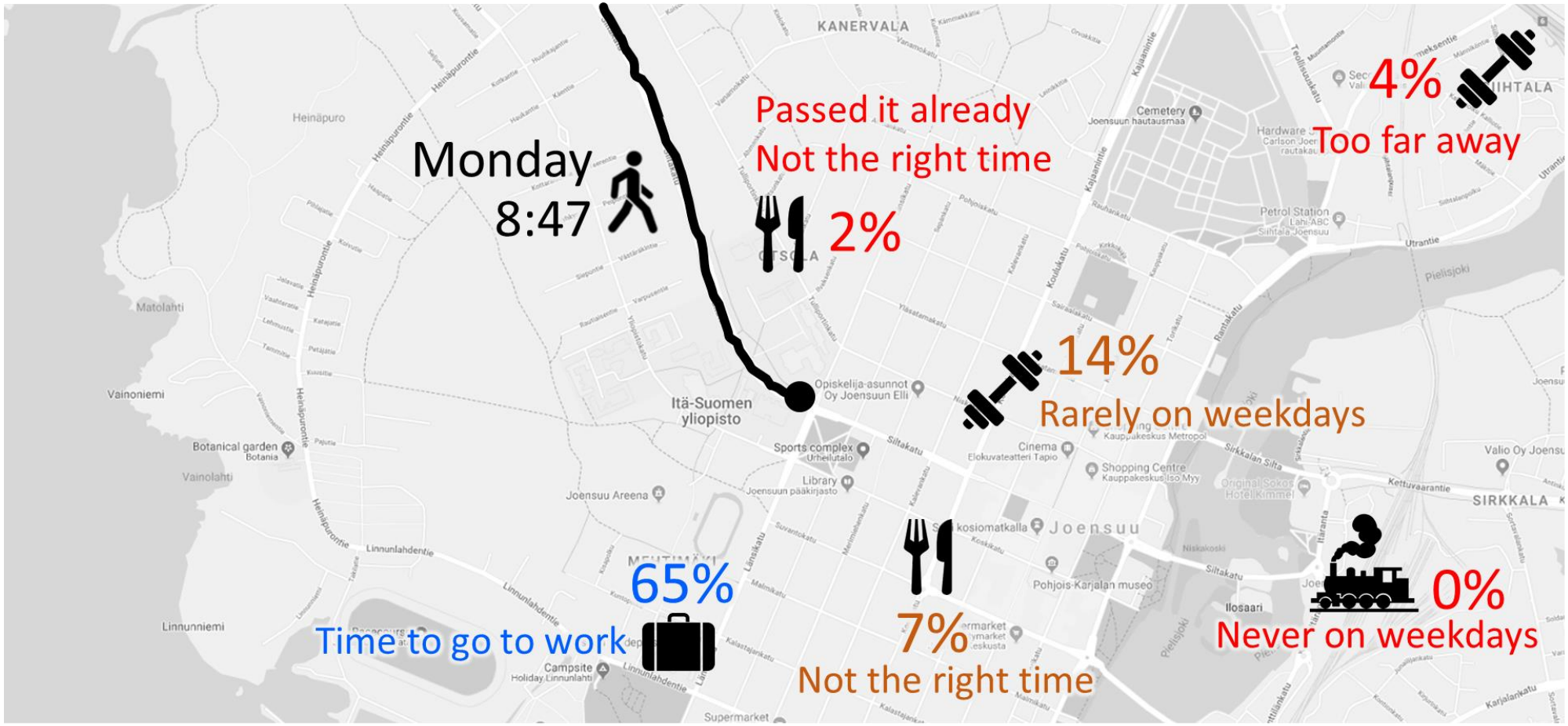
N 61.19 E 29.01

R. Mariescu-Istodor

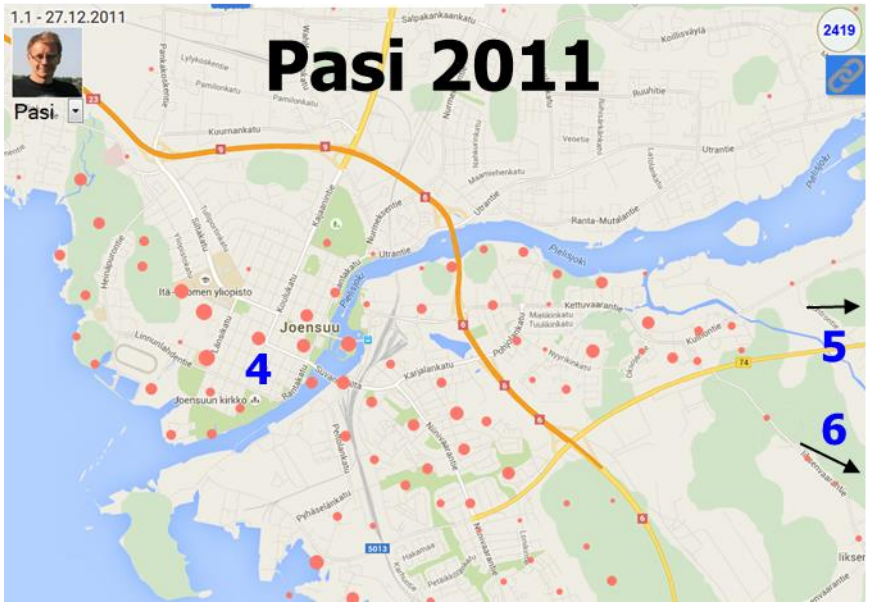
Destination prediction

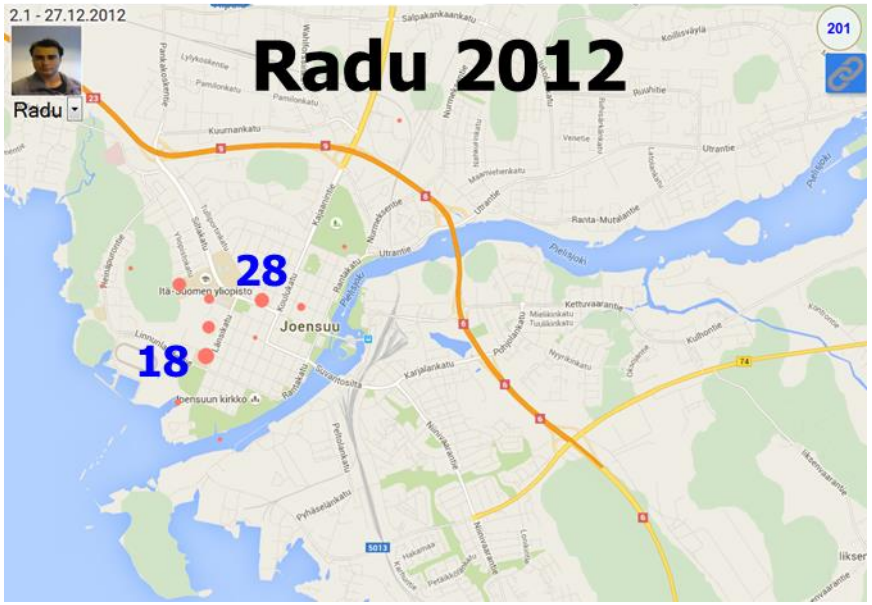
Radu Mariescu-Istodor, Roxana Ungureanu and Pasi Fränti,
"Real-time destination prediction for mobile users." *LBS* **2019**

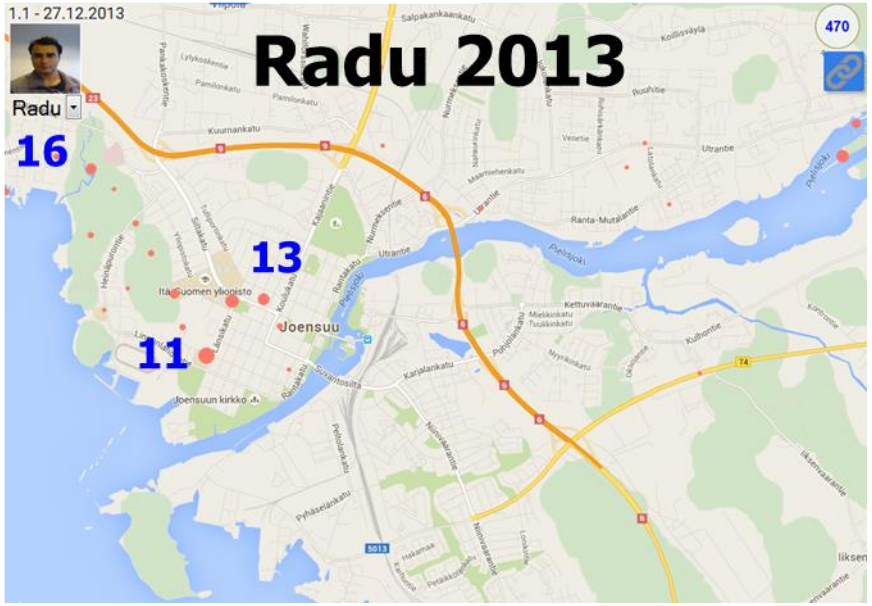
R. Ungureanu



Same user or not?





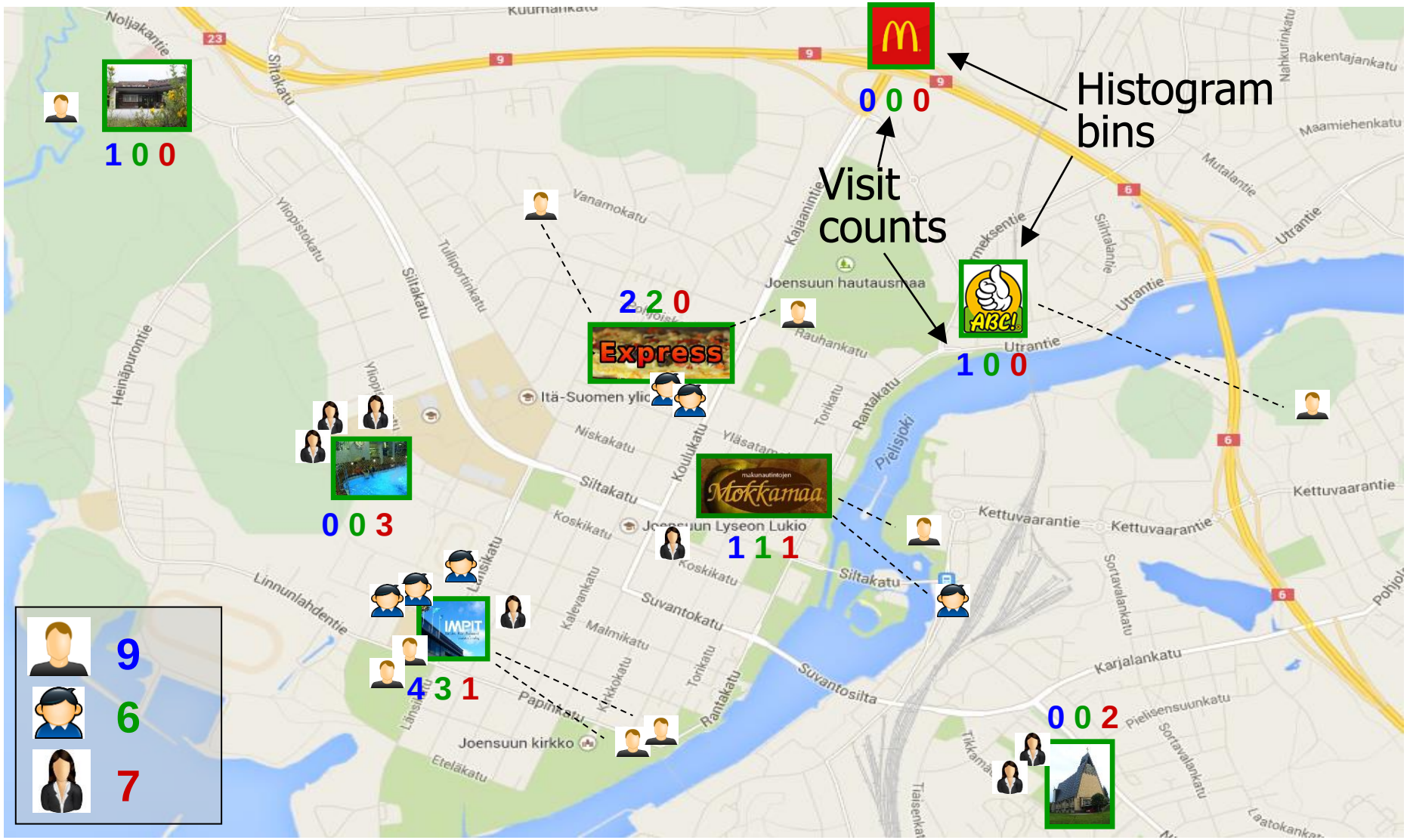


K. Waga

Location history

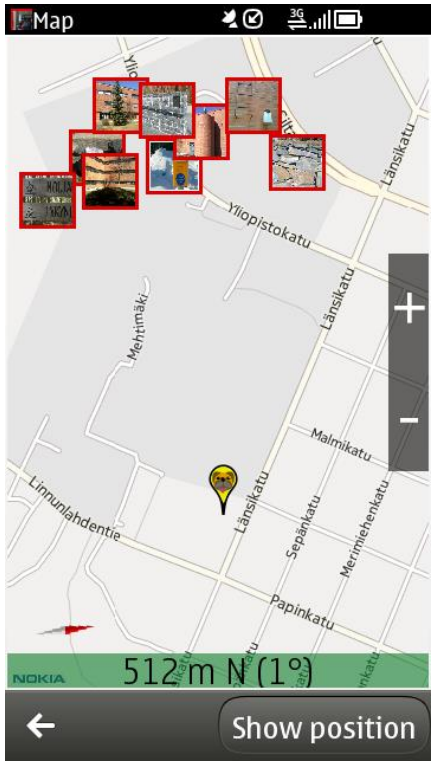
Fränti, Mariescu-Istodor and Waga, "Similarity of mobile users based on sparse location history", *ICAISC* **2018**

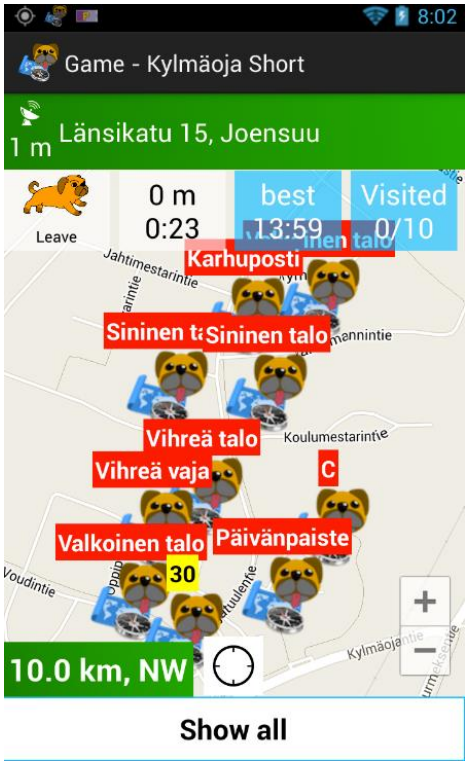
R. Mariescu-Istodor



O-Mopsi orienteering game

P. Fränti, R. Mariescu-Istodor and L. Sengupta,
O-Mopsi: mobile orienteering game for sightseeing, exercising and education
ACM Trans. on Multimedia, Computing, Communications, and Applications, **2017**





Promoting

Call for papers

Applied Computing and Intelligence

https://aimspress.com/journal/aci

Highlights of the journal :

- Scope on AI and computing
- Open access
- No publication charges
- Fast review process

A good paper has:

- Clearly stated contribution
- Focused topic
- New insight gained
- Easy to read

AMERICAN INSTITUTE OF MATHEMATICAL SCIENCES