

14.11.2017



# Thesis topics

**Pasi Fränti**

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



# **Web mining**

# Automatic creation of web page

Input: Mopsi services

Output: Simple web page

Motive:

- Mopsi data easy to create.
- Automatic. Decent outlook.

Data:

- Sample photos
- Address
- Map
- Description

Closest existing solutions:

- CMS: Drupal, wordpress
- <https://joensuunyt.wordpress.com/>
- <https://vintagejoensuu.wordpress.com/>

Started JS

Title : Pizza Master  
Keywords : cheap, student, best, pizza  
Description : Pizza and kebab restaurant  
Address : Niskakatu 2, 80100 Joensuu

Photos :



## Pizza Master

Pizza and kebab restaurant

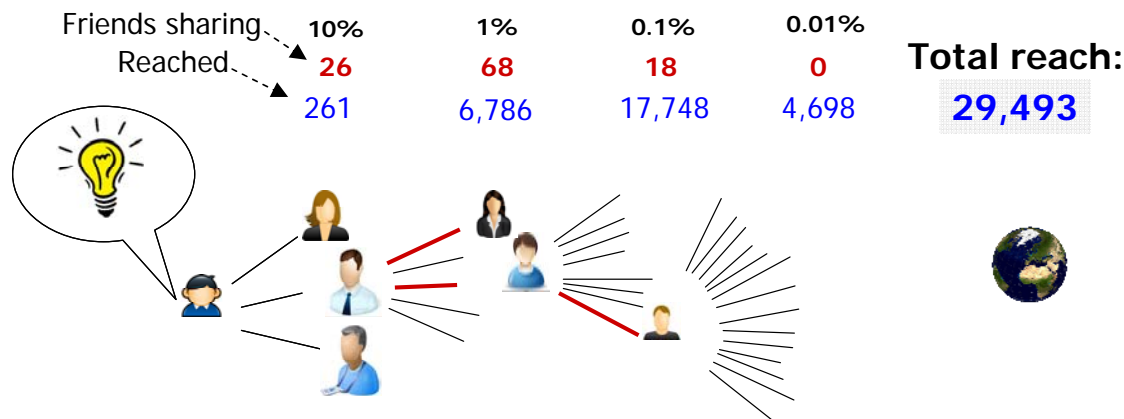


- Cheap
- Student
- Best
- Pizza



Niskakatu 2, 80100 Joensuu

# Ad hoc and social networks analysis



## 1. Ad hoc -verkoston luominen eri informaatiosta.

Käyttäjien välisestä kommunikoinnista (on relaatio), tai heidän käyttäytymisestään (ovat usein samaan aikaan samassa paikassa). Pelkästään sijainnista ja aikainformaatiosta voi päätellä jotain jos dataa tarpeeksi.

## 2. Verkoston hyödyntäminen markkinoinnissa.

Informaatioteoreettisempi näkökulma. Mikä on kustannus että saadaan informaatio henkilöltä A henkilölle B? Voidaan laskea myös hyötyfunktio, jolloin voidaan mallintaa ihmisten saavutettavuutta verkoston kautta todennäköisyysmalleilla.

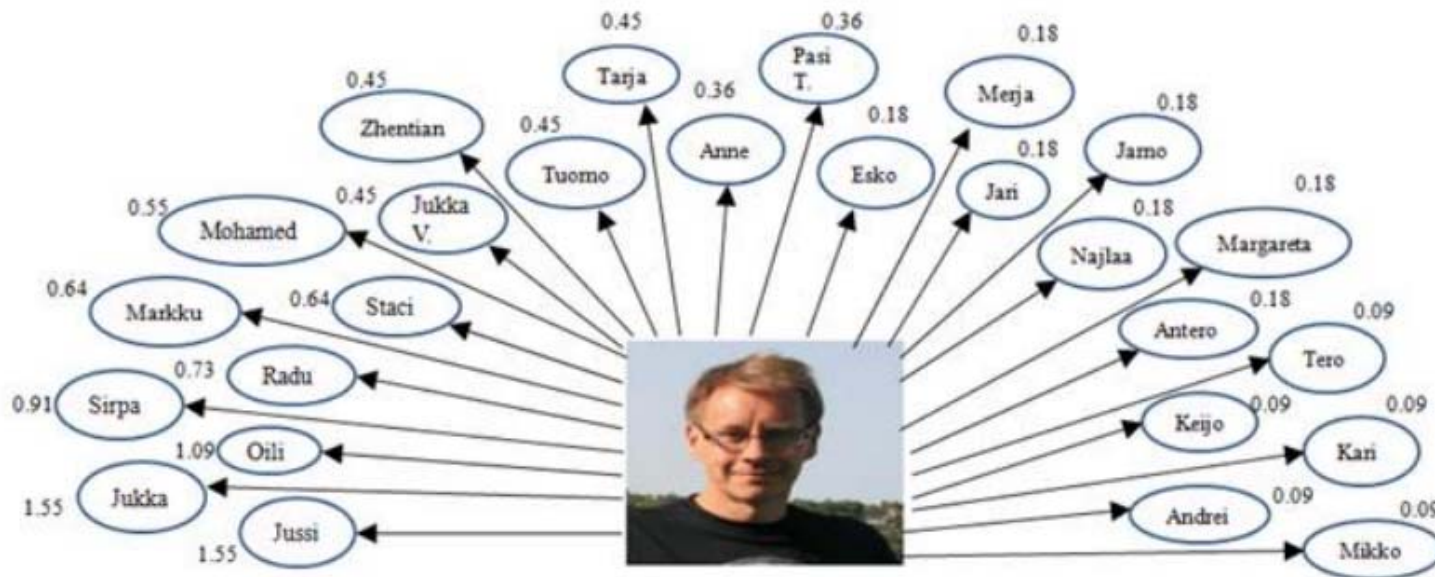
**3. Sijainnin hyödyntäminen verkostossa:** Sijaintitiedon muuttumisen analysointi. Ensin klusteroidaan paikkakeskittymät (joissa ihmiset kulkee). Sitten segmentoidaan henkilön data ajan funktiona. Tarkastellaan muutoksia. Päätellään milloin henkilö "vaihtaa paikkaa". Informointi kavereille kamujen liikkeistä. Esim. jos Pekka saapuu Joensuuhun (yksi klusteri) sillä on merkitystä Pasille. Mutta jos Joensuussa asuva käy naapurikunnassa ja tulee kohta takaisin, ei välttämättä. Jos Joensuussa asuva saapuu tiedepuistoon (Joensuun sisällä oleva klusteri), niin voi olla jo relevanttia.

**4. Verkoston analysointi:** linkkien vahvuus, klusterointi, solmujen etäisyydet. Ei sen kummempia sovellustavoitteita. Mutta jos perusanalyysit osataan tehdä, näitä on helppo soveltaa ja hyödyntää monella tavalla joita ei vielä ole osannut edes ajatella.

Kantava ajatus, että täytyy konkreettisesti liittyä tähän sovellukseen. Muuten jää helposti irralliseksi ja lähtee meneen johonkin ehkä mielenkiintoiseen mutta epärelevanttiin suuntaan.

# Social network in FB and Twitter

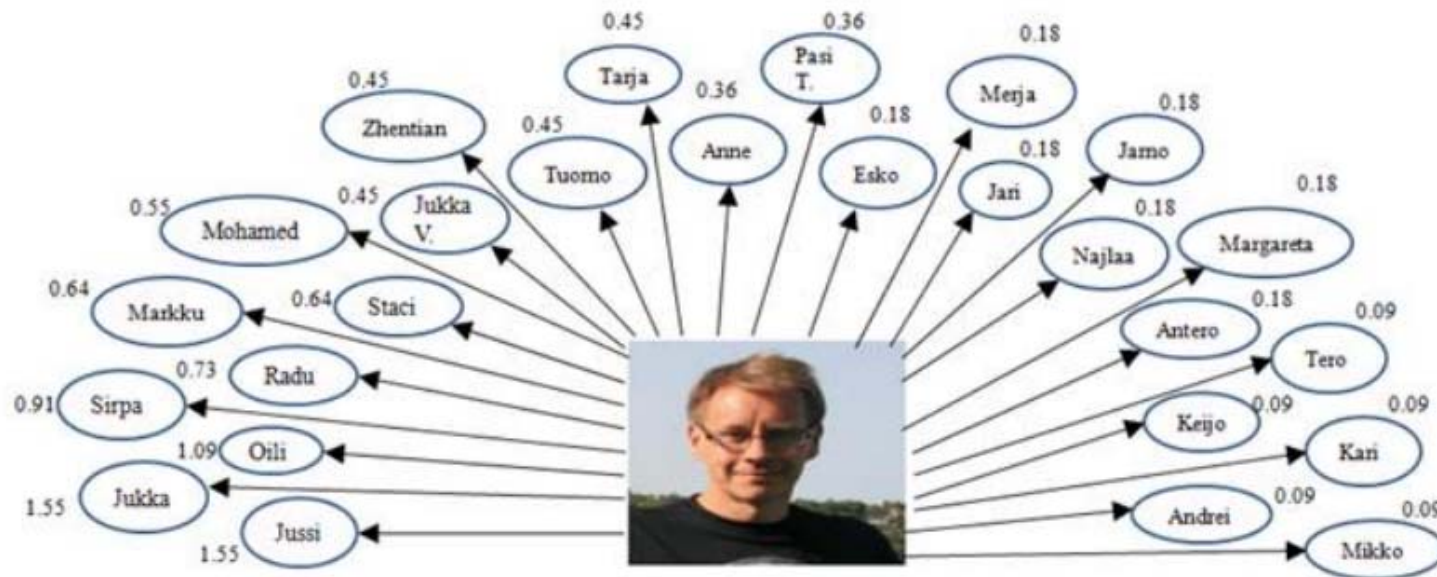
- Build Facebook/Twitter API to fetch data
- Analyze network structure (e.g. community detection)
- Analysis of user behavior, similarity
- Alternative focus on sociology theme combining user self-image and user influence.



Pasi's influence = 11.9

# Data exchange Web ↔ Facebook

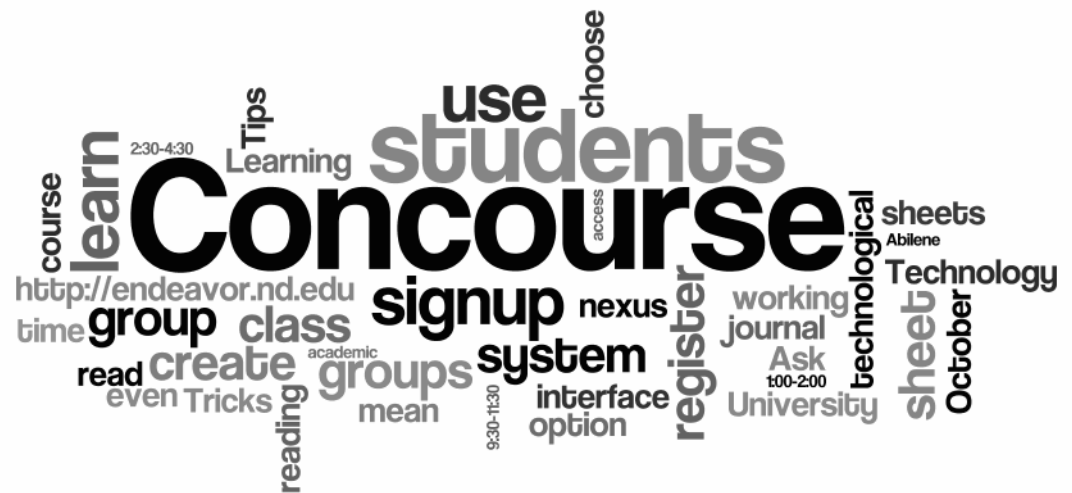
- Convert Web page to FB publications
- Convert FB content to Web page (photo albums, notes)
- Requires web mining (DOM tree, data extraction, NLP)
- Web programming and FB API skills required
- Data analysis




Pasi's influence = 11.9

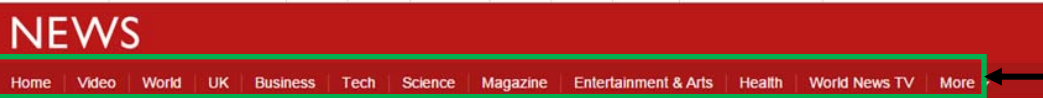
# Word cloud from web page

- Analyze given web page
- Extract relevant keywords
- Build **word cloud** from the keywords
- Optional: clustering of the keywords
- Possible applications:
  - Person homepage
  - Facebook publications



# Language-independent content extraction

Logo image → 

Navigation bar → 

Title → **Raspberry Pi 3 adds wi-fi and Bluetooth**

Keywords → **The Raspberry Pi**

Text → 

The title was formerly held by the Amstrad PCW which is believed to have sold a total of eight million units.

Sales of the Raspberry Pi will surpass that figure this month, said the Raspberry Pi project founder Eben Upton.



Technology

**Raspberry Pi 3 adds wi-fi and Bluetooth**

2 hours ago | Technology



UK astronaut Tim Peake took a Raspberry Pi to the International Space Station

**The Raspberry Pi** has become the most popular British computer yet made.

The title was formerly held by the Amstrad PCW which is believed to have sold a total of eight million units.

Sales of the Raspberry Pi will surpass that figure this month, said the Raspberry Pi project founder Eben Upton.

## Top Stories

Oscars 2016: DiCaprio finally wins

1 hour ago

**UN to expand Syria aid as truce holds**

30 minutes ago

**Pakistan hangs killer of state governor**

4 hours ago

## Features & Analysis



**He said yes!**

Eight women who proposed to their partners



Navigation bar

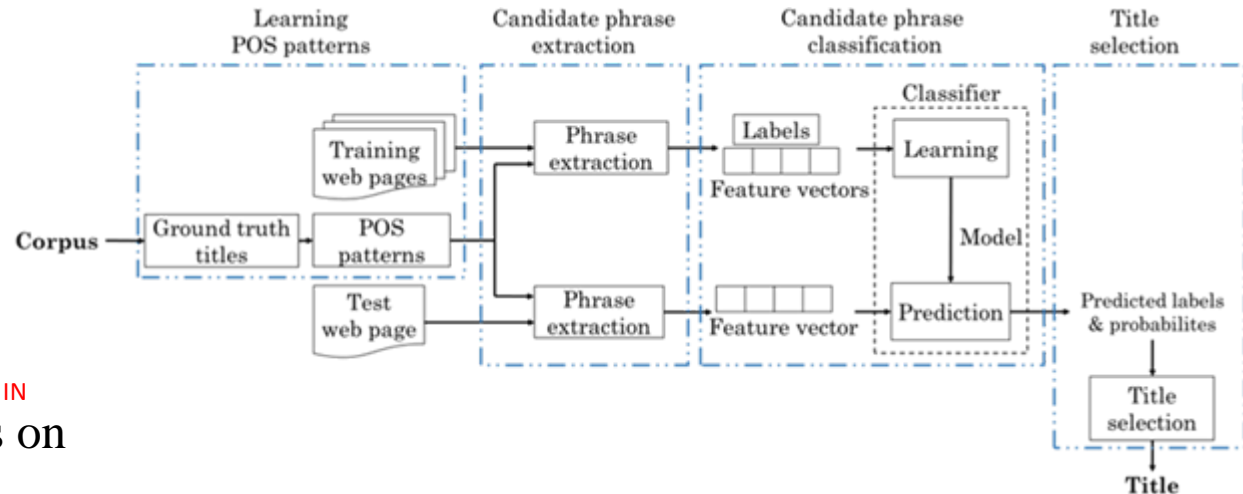
Images



# NLP tools for detecting tweeting Bots

4-month project available!

<http://cs.uef.fi/sipu/OpenPositions-2.pdf>



NNP  
Navigation

VBG NNP VB PRP IN  
Feeling Social? Find us on

NNP  
Facebook

NNP NNP NNP NNP NNPS NN  
Sydney Waterfront Restaurant Restaurant Milsons Point

NNP NNP VBZ DT JJ NNP NN NN  
Aqua Dining offers a quintessential Sydney dining experience  
IN JJ NN NNS WDT NN IN NNP NNP IN DT  
with unrivalled harbour views that sweep from Luna Park to the  
NN JJ NNP NNP NNP CC DT NNP NNP  
world famous Sydney Harbour Bridge and the Sydney Opera  
NNP  
House.

NNP=Proper noun, singular  
NNPS=Proper noun, plural  
NN=Noun, singular or mass  
VBG=Verb, gerund  
VB=Verb, base form  
PRP=Personal pronoun  
DT=Determiner  
CC=Coordinating conjunction  
JJ=Adjective

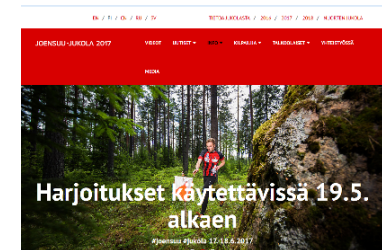
# Correlation-based word similarity

Input: Two words

Output: Similarity in range 0..1

Method: use web searches to find out co-occurrences of the words. Estimate similarity based on this and their total frequency w/o any language information.

Example: **Uutinen**, **news**  
High similarity because same meaning despite different language.



# Detecting clusters from text phrases

**Input:** Given a set of geo-tagged user photos in (1) Mopsi, (2) web search.

**Goal:** Detect potential cluster.

**Approach:** Cluster first by location. Then find out whether there is dominant cluster based on their descriptions. For example, can we conclude what is in the following location?

## Descriptions:

Kauppa valintatalo  
Kahvila heinosen leipomo  
Valintatalo kauppa  
Kauppa  
Kahvila Heinosen leipomo  
Äijähetki salaatti - box fit!!!  
Valintatalo ruokakauppa  
Valintatalo  
Nyt välähti  
Market  
Scientific Writing w 10l milk.. :)  
Keltainen talo  
Valintatalo

Started  
DH+RM



# Clustering

# Density-based clustering

Clustering with weighted centroids:

- Each centroid has weight  $[0..1]$
- Cost is weighted distance
- Valuable centroids attract points far away
- All weights sum=1 (or some constant)
- Resource allocation problem

Algorithm:

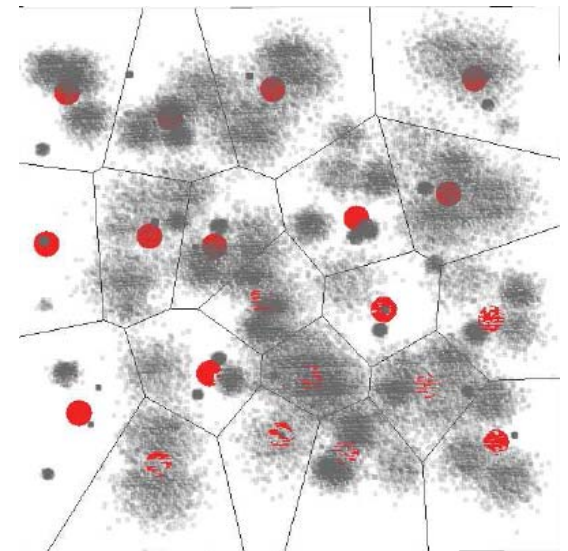
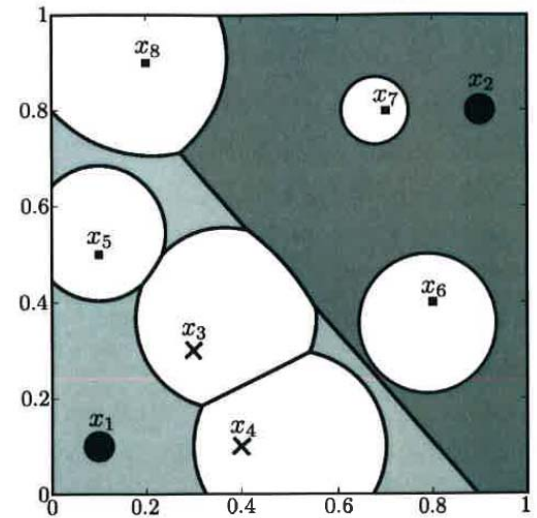
- K-means / RS algorithm
- Open question: how to optimize weights?

Experimenting:

- CS-software with modules (c-lang)
- Aim at solving Birch3 dataset and such

Alternative:

- Graph-based density estimation



# K-means properties

- Systematic study on K-means with artificial datasets varying
  - Size
  - Number of clusters
  - Dimensionality
- Properties of the datasets (A, G2, DIM,...)
  - Neighborhood size
  - Clusterness
- Systematic study when K-means work and when not
- Number of iterations
- Stability
- Dimensionality properties of G2
- Extension to other cost functions beyond TSE

**Completed**

# Balance k-means clustering

M.I. Malinen and P. Fränti, "Balanced K-means for clustering", *Joint Int. Workshop on Structural, Syntactic, and Statistical Pattern Recognition (S+SSPR 2014)*, LNCS 8621, 35-44, Joensuu, Finland, August 2014.

Mikko's S+SSPR paper as starting point

Alternatives for partition step:

- As pairing problem using Hungarian algorithm:  $O(n^3)$
- Network flow problem (for balancing)
- Any heuristics where centroids have size constraint

Balance-driven cost function

- $\lambda=0$  is k-means;  $\lambda = \infty$  is balance
- $\lambda$  is increasing until converged to balance solution

$$\sum \|x_i - c_j\|^2 + n_j \cdot \lambda$$

# Stability of clustering algorithms

- How much results of clustering changes if randomness is added to data
- How much if added to the clustering process
- For example comparing result of full set and random subsets (20%):
  - Stable algorithm provides same result, unstable different.
  - Such experiments are probably done for K-means only.
- Requires systematic testing, re-compiling software, and extensive testing



# Dynamic Random Swap

- Variant that solves also number of clusters
- Three centroid operations:
  - Swap
  - Removal
  - Addition (instead of split)
- Proper stopping criterion
- Cost function should also solve k. Possibilities:
  - Q. Zhao and P. Fränti, "WB-index: a sum-of-squares based index for cluster validity", *Data & Knowledge Engineering*, 92, 77-89, July 2014
- Start variant is reported here:
  - I. Kärkkäinen and P. Fränti, "Dynamic local search for clustering with unknown number of clusters", Int. Conf. on Pattern Recognition (ICPR'02), Québec, Canada, vol. 2, 240-243, August 2002.
  - <http://cs.uef.fi/sipu/pub/DynamicLocalSearch-ICPR2002.pdf>

# Clustering users for recommendation

## Clustering users

- Based on social network (common neighbors)
- Based on similar interests (likes in FB)

## Recommendation:

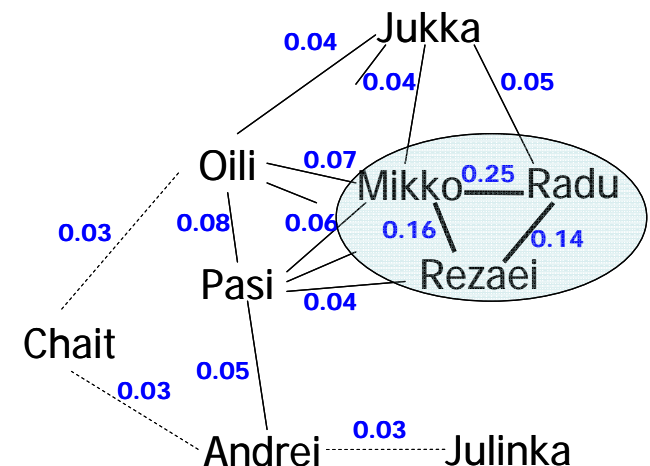
- $\text{Relevance} = \text{Influence of user} * \text{Relevance of POI}$
- $\text{Influence} = \text{Activity} + \text{dominance}$

## Approach A:

- Group users together
- Same recommendation for all users

## Approach B:

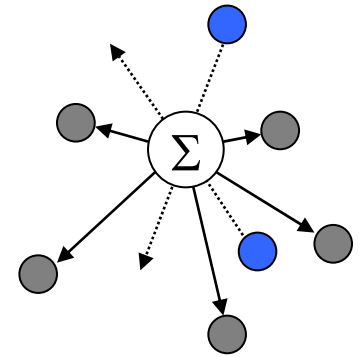
- Individual recommendations
- Merge recommendations



# Clustering-based classifier

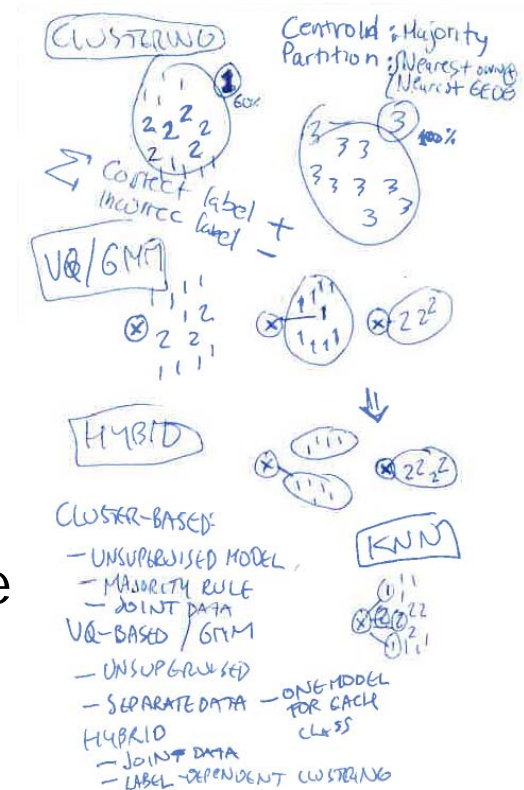
Traditional centroid-based models

- Optimize separate model for each class
- Adopt from universal background model
- Vector quantization (clustering-based) or GMM
- New: generate joint model



Joint optimization (new)

- Shared prototypes
- Training vectors join to nearest cluster
- Label of cluster prototype by:
  - (a) Predefined label
  - (b) Adaptively by majority rule
- Vectors with same label attract prototype
- Vectors with different label resist prototype



# Optimizing classifier weights by GA

- Good understanding of basic classifiers
- Interest in machine learning
- C/Java programming

# Clustering categorical data

	VSC-896	ABC-123	TDK-111
Model	VW	Opel	Hyundai
Color	Silver	Red	Blue
Country	Germany	Germany	Korea
Owner	Pasi	Aku	Pasi

- Connection via sharing attribute:
  - VSC-896 and TDK-111 have same owner
  - VSC-896 and ABC-123 made in same country
- Connection can be exact (=same) or inexact (European, Asian)
- Correlation of attributes connections correlate
- Euclidean:
  - Only nearby distances (KNN or XNN)
  - Connection with strength  $1/d$
- Clustering:
  - Cluster goodness = maximal (average) similarity:  $\text{SUM}(1/d) = \text{max!}$
  - $\text{MSE} = \text{SUM}_{ij}(d^2)/n = \text{min!}$
  - $\text{MSE} = \text{SUM}_{ij}(1/d^2)/n = \text{max!}$

**TSP**

# TSP algorithms

## Suitable also for DAA projects

### 1. Christofides

- Complete implementation using existing components

### 2. Pre-Christofides

- The simple 1-approximation variant using tree traversal

### 3. Kruskal TSP

- Build TSP similar to Kruskal
- Allow only adding links to the end of chains

### 4. MST → TSP

- Follows the spirit of Christofides
- Instead of adding more links, detect knots and leafs
- Re-connect leafs and knots to remove branches

### 5. Local search

- Use 2-opt, node swap, link swap operations
- Allow branches in the intermediate solutions

# O-Mopsi and games





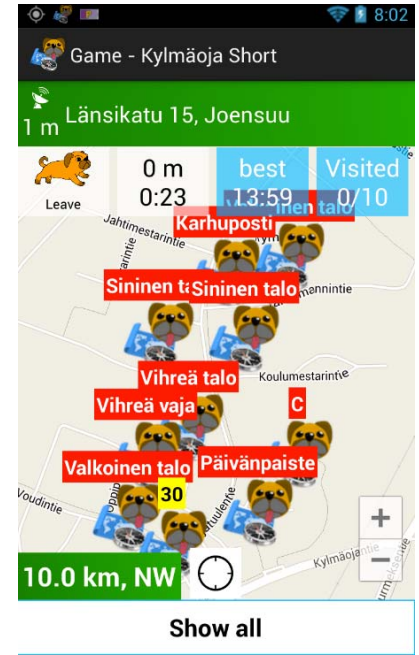
# O-Mopsi Android

## Goal:

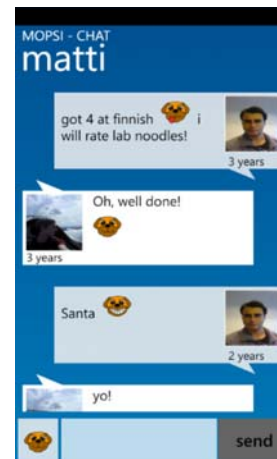
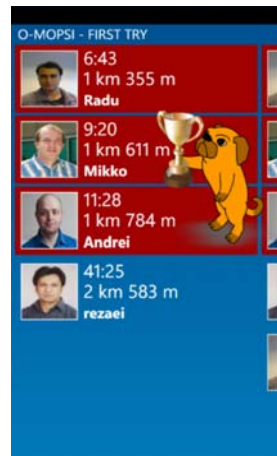
- New features for O-Mopsi game
- Chat, feedback, top score
- Content creation tools

## Requirements:

- Strong programming experience
- Ready to learn Android
- Might require web-programming



No.	Photo	Player name	Finished
1		Jukka	35
2		Pasi	30
3		Radu	20
4		Mikko	7
5		LahariS	6



# Educational applications

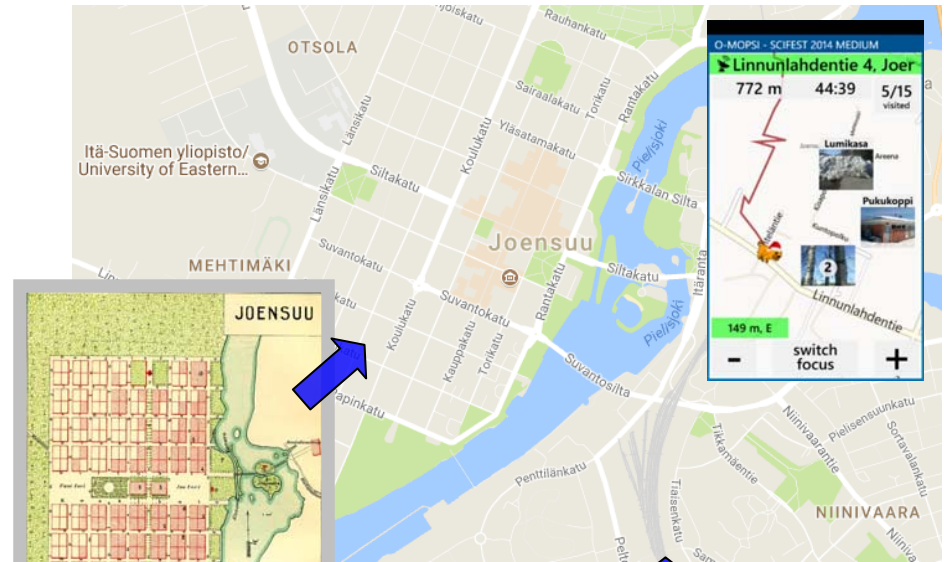
## Tools for adding material:

- Additional map layers
- Questions & answers
- More info about target

## Educational games?

YOU GOT 100% AVG SCORE 24% Even Tina Turner agrees: you're simply the best! (like) SCORE 50/50 TIMER 04:55

TRY AGAIN SHARE RESULTS QUIZ STATS PLAY ANOTHER



# Educational content

- Name of object (trees, flowers)
- Background story

Outdoor museums

Geography

Botanical gardens

Biology

Architecture

City history

Exhibitions



Aspen



Lime tree



Maple



Oak



Elm



Thuja



Spruce



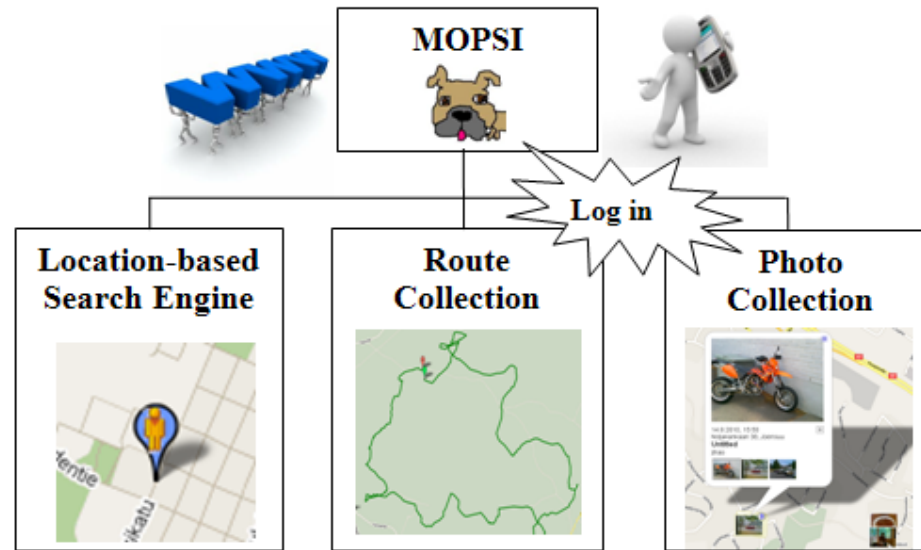
Birch



Rowan

**Mopsi**

# Mopsi overview



## 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

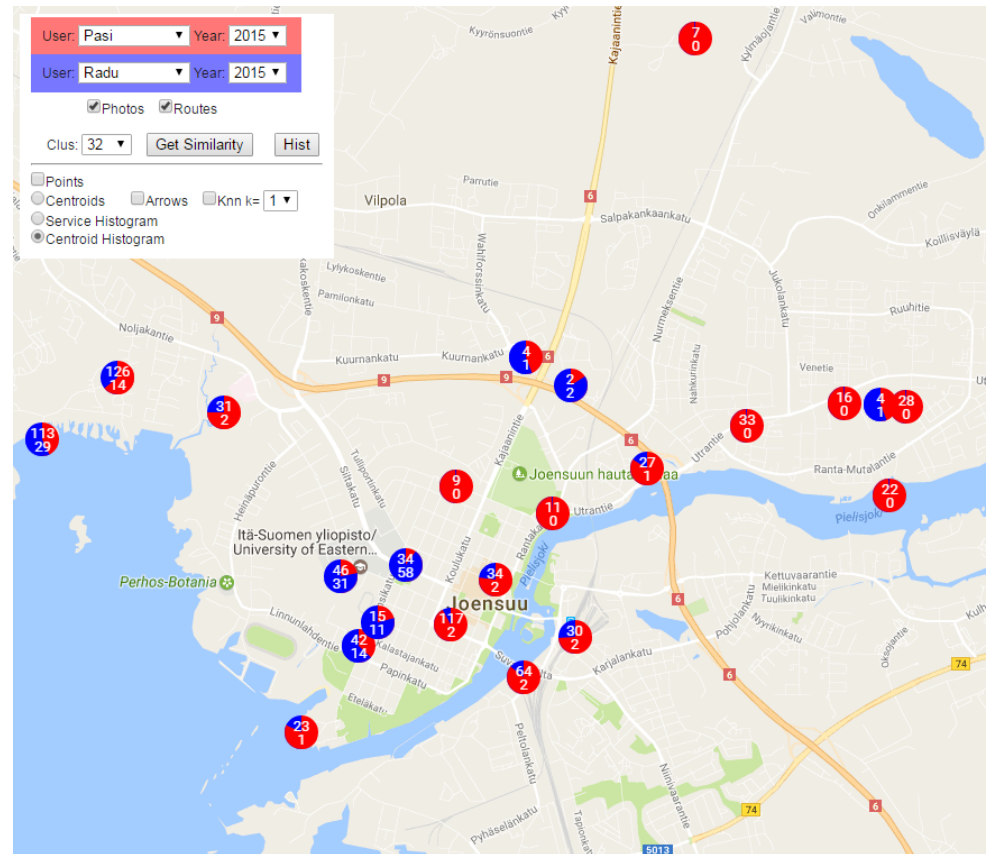
## Location-based Game



- Orienteering
- Killer-game



# User similarity





- Studying different methods based on user data
- Profile, activity, location history, photo descriptions
- Requires machine learning skills such as
  - histogram comparison
  - clustering
  - model adaptation



# Clustering routes

**MOPSI** Radu's routes 21.5.2015 - 20.5.2016 Pasi Downloads O-Mopsi Tools






search using a keyword Recommend  


**Routes** Photos Both

Most recent Week Month **Year** All Select dates


Show points

-  11878 km
-  1650 km
-  361 km
-  257 km
-  13.7 km


**19.5.2016**

-  **Route 1:** 16:40 - 18:15 41 km


**13.5.2016**

-  **Route 2:** 16:51 - 18:38 41 km



**7.5.2016**

-  **Route 3:** 13:23 - 18:05 7 km 595 m


**6.5.2016**

-  **Route 4:** 13:14 - 13:37 2 km 002 m

**23.4.2016**

-  **Route 5:** 09:35 - 10:27 8 km 025 m
-  **Route 6:** 10:31 - 11:55 16.9 km

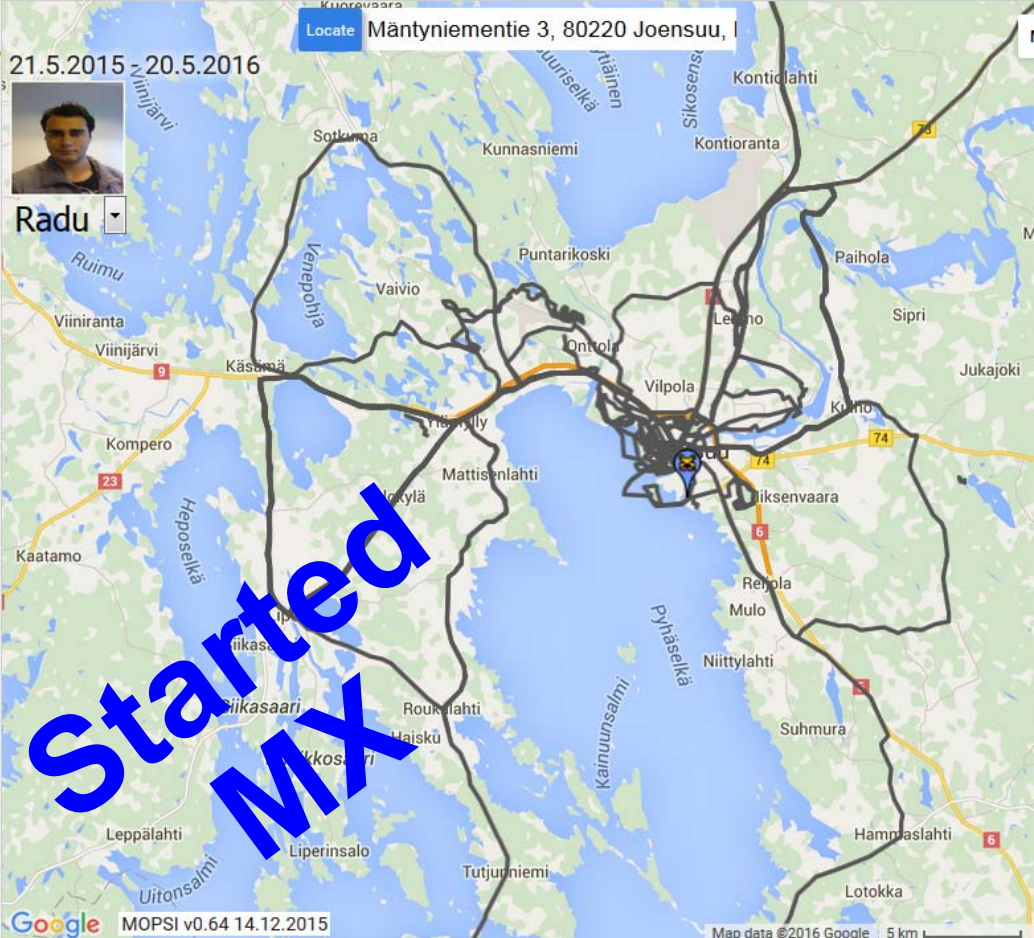
**21.4.2016**

-  **Route 7:** 14:31 - 16:02 4 km 978 m

**Radu**

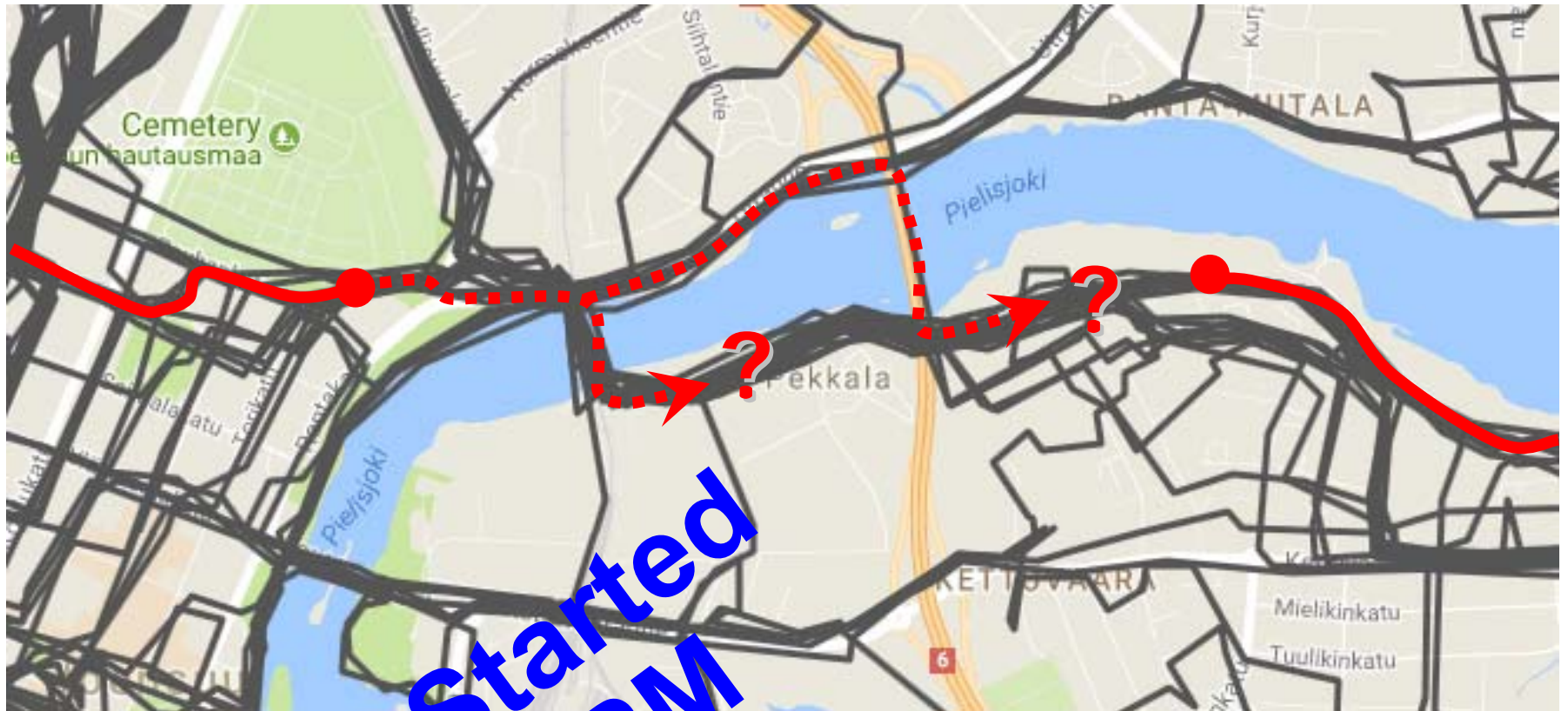
Mäntyniementie 3, 80220 Joensuu, I

21.5.2015 20.5.2016



Google MOPSI v0.64 14.12.2015 Map data ©2016 Google 5 km

# Route prediction



Started  
RM



# Organizing route collection for efficient retrieval

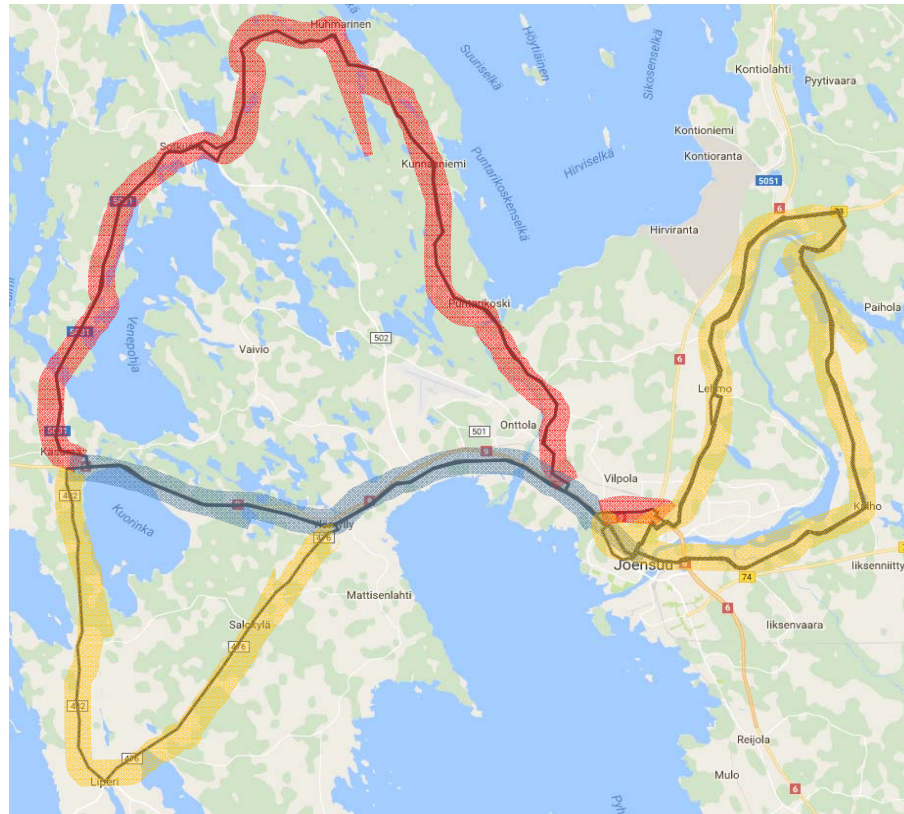
<http://cs.uef.fi/sipu/pub/Gsearch.ppt>



R. Mariescu-Istodor and P. Fränti, "Gesture input for GPS route search", *Joint Int. Workshop on Structural, Syntactic, and Statistical Pattern Recognition (S+SSPR 2016)*, Merida, Mexico, LNCS 10029, 439-449, November 2016.

# GPS route search via segmentation

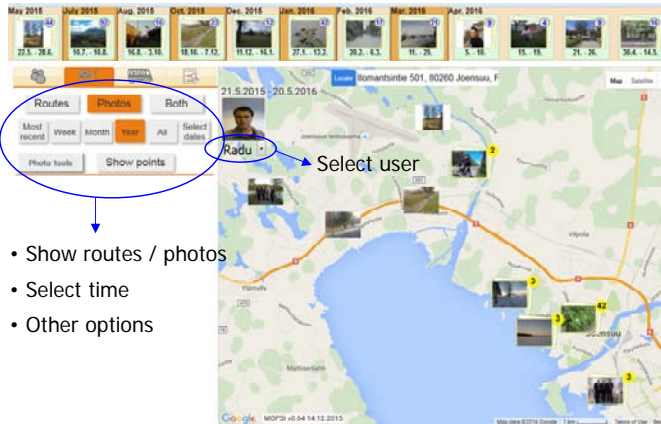
- Segment the on-screen routes using Road Network
- Clicking a segment limits to routes passing through the segment
- Clicking a second segment continues to limit the screen, etc.



# Optimizing health care services

3-years project starting 1.1.2018

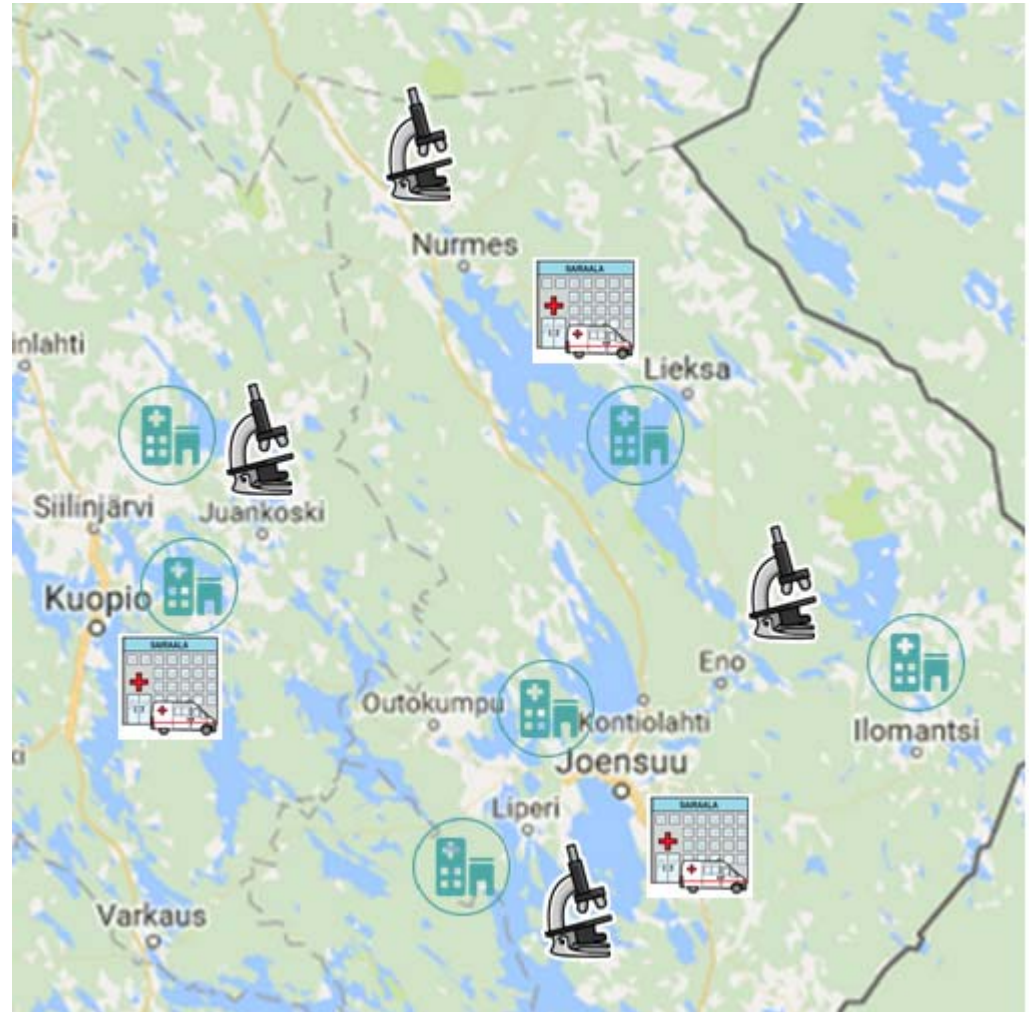
<http://cs.uef.fi/sipu/OpenPositions-1.pdf>



- Show routes / photos
- Select time
- Other options

## Main goals:

- Developing methods for location-based applications
- Apply the methods for multi-objective optimization (cost, location, travel time)
- Interactive user-interface on maps



# Students on Map

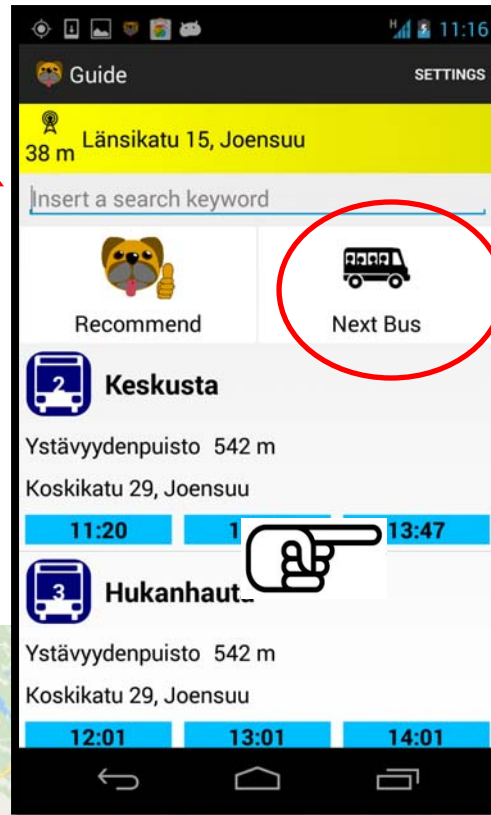


# Mobile bus time table systems

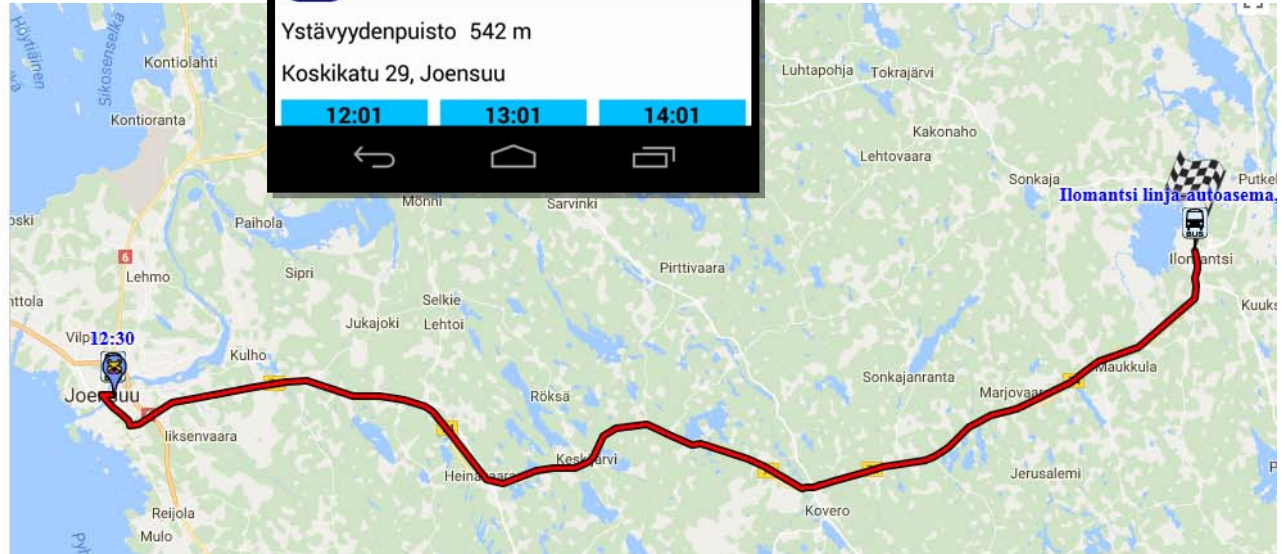
Started  
GR

Location

One button  
system



11:36	4	Joensuun pääkirjasto I	11:36	
11:39	31	9	Siltakatu B L	11:39
11:50	9811	45	Sairaala PKKS E	11:50
12:00	9541	55	Outokumpu linja-autoasema	12:00
12:00	null	55	Kuopio linja-autoasema	12:00
12:05	9821	60	Lieksa linja-autoasema	12:05
12:20	9140		Kiitelysvara koulu, P	12:20
12:30	9120		Ilomantsi linja-autoasema	12:30
12:30	9623		Polvijärvi linja-autoasema	12:30



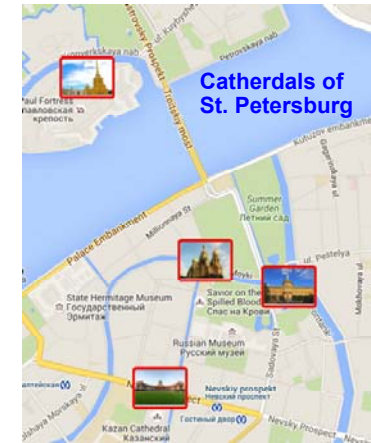
# Web crawling for collecting content

## Game targets:

- Outdoor landmarks
- Objects in park
- User's own travel pictures

## Content-creation:

- Manually maintained (Mopsi database)
- External geo-tagged databases
- Collecting by web crawling



**Started  
KN**

## Goal:

Creaste tool for systematic collection of material by crawling for geo-tagged pictures

# Location-aware recommendation

Given location, recommend  
Relevant places around

- Vilku Kahvio**  
Pamilonkatu 33, Joensuu  
740 m
- Kuurnankulma**  
Paukkajantie 2-4, Joensuu  
762 m
- Kahvila Heinosen leipomo**  
Hiiskoskentie 13, Joensuu  
306 m
- La Dolce Vita**  
Kuurnankatu 6, Joensuu  
625 m
- Kahvila Huili & Javerstok grilli**  
Kuurnankatu 14, Joensuu  
420 m
- Skarppi - ylioppilaskunnan sauna**  
Kaislaku 10, Joensuu  
758 m

## Vilku kahvio

Location	0.55
Time	1.00
Relevance	0.84
User network	0.57 ★★☆☆
<b>Total:</b>	<b>2.95</b>

	All		
Keywords	Nearby	Recently	Total
kahvila	0.32	0.00	0.27
<b>Total (Max)</b>	<b>0.32</b>	<b>0.00</b>	<b>0.27</b>
<b>Total (Av)</b>	<b>0.32</b>	<b>0.00</b>	<b>0.27</b>

## Kuurnankulma

Location	0.60
Time	1.00
Relevance	0.32
User network	1.00 ★★★★★
<b>Total:</b>	<b>2.92</b>

	All		
Keywords	Nearby	Recently	Total
lounas	0.02	0.00	0.02
<b>Total (Max)</b>	<b>0.02</b>	<b>0.00</b>	<b>0.02</b>
<b>Total (Av)</b>	<b>0.02</b>	<b>0.00</b>	<b>0.02</b>

**Vilku kahvio**  
740 m

**Kuurnankulma**  
762 m

**Heinosen leipomo**  
306 m

## Heinosen leipomo

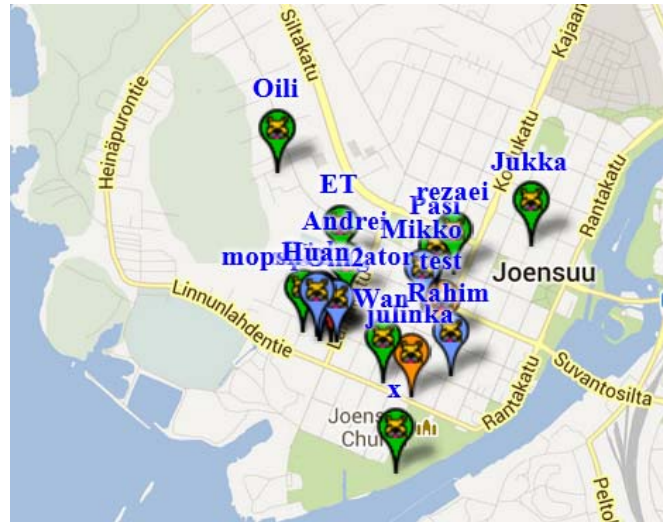
Location	0.95
Time	1.00
Relevance	0.84
User network	0.00
<b>Total:</b>	<b>2.79</b>

	All		
Keywords	Nearby	Recently	Total
kahvila	0.32	0.00	0.27
<b>Total (Max)</b>	<b>0.32</b>	<b>0.00</b>	<b>0.27</b>
<b>Total (Av)</b>	<b>0.32</b>	<b>0.00</b>	<b>0.27</b>

# Lenkille.com

Kimppalenkki (miksei myös kyytien) suunnitteluun ja koordintointiin tarkoitettu työväline. Hyödyntää käyttäjien sijaintia (Mopsi). Käyttäjä voi luoda uuden tapahtuman (event) tai liittyä olemassa olevaan. Järjestelmä tunnistaa osallistumiset automaattisesti ja informoi osallistujia. Piirteitä FB events, nimenhuuto.com ja kimppakyyti-sovelluksista.

Started  
AS



## Planned events:

- Pekka: Lauantai 10.00 Areena Lenkki 18-20 km
- Olli: Tiistai 18.00 Tiedepuisto Juoksu 10 km
- Radu: Keskiviikko 16.00 Vesikko Hiihto 12 km



# Publications

## Routes

1. R. Mariescu-Istodor and P. Fränti, "Grid-based method for GPS route analysis for retrieval", *ACM Trans. on Spatial Algorithms and Systems*, 3 (3), 8:1-28, September, 2017.
2. R. Mariescu-Istodor and P. Fränti, "Gesture input for GPS route search", *Joint Int. Workshop on Structural, Syntactic, and Statistical Pattern Recognition (S+SSPR 2016)*, Merida, Mexico, LNCS 10029, 439-449, November 2016.
3. S. Sieranoja, T. Kinnunen and P. Fränti, "GPS trajectory biometrics: from where you were to how you move", *Joint Int. Workshop on Structural, Syntactic, and Statistical Pattern Recognition (S+SSPR 2016)*, Merida, Mexico, LNCS 10029, 450-460, November 2016.
4. R. Mariescu-Istodor, A. Tabarcea, R. Saeidi and P. Fränti, "Low complexity spatial similarity of GPS trajectories", *Int. Conf. on Web Information Systems & Technologies (WEBIST'14)*, Barcelona, Spain, April 2014.
5. K. Waga, A. Tabarcea, R. Mariescu-Istodor and P. Fränti, "Real time access to multiple GPS tracks", *Int. Conf. on Web Information Systems & Technologies (WEBIST'13)*, Aachen, Germany, 293-299, May 2013.
6. K. Waga, A. Tabarcea, M. Chen and P. Fränti, "Detecting movement type by route segmentation and classification", *IEEE Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom'12)*, Pittsburgh, USA, 2012.
7. M. Chen, M. Xu and P. Fränti, "A fast  $O(N)$  multi-resolution polygonal approximation algorithm for GPS trajectory simplification", *IEEE Trans. Image Processing*, 21 (5), 2770-2785, May 2012.
8. M. Chen, M. Xu and P. Fränti, "Compression of GPS trajectories using optimized approximation", *IEEE Int. Conf. on Pattern Recognition (ICPR'12)*, Tsukuba City, Japan, 3180-3183, November 2012.
9. K. Waga, A. Tabarcea, R. Mariescu-Istodor and P. Fränti, "System for real time storage, retrieval and visualization of GPS tracks", *Int. Conf. System Theory, Control and Computing (ICSTCC)*, Sinai, Romania, Vol. 2, October 2012.
10. M. Chen, M. Xu and P. Fränti, "Compression of GPS trajectories", *IEEE Int. Conf. on Data Compression Conference (DCC'12)*, Snowbird, Utah, 62-71, April 2012.

# Publications

## Web Mining

1. N. Gali, R. Măriescu-Istodor and P. Fränti, "Using linguistic features to automatically extract web page title", *Expert Systems with Applications*, 79, 296-312, 2017.
2. N. Gali, R. Măriescu-Istodor and P. Fränti, "Similarity measures for title matching", *IAPR Int. Conf. on Pattern Recognition, (ICPR'16)*, Cancun, Mexico, 1549-1554, December 2016.
3. N. Gali and P. Fränti, "Content-based title extraction from web page" , *Int. Conf. on Web Information Systems and Technologies (WEBIST 2016)*, Rome, Italy, vol. 2, 204-210, April 2016.
4. M. Rezaei, N. Gali, and P. Fränti, "CIRank:a method for keyword extraction from web pages using clustering and distribution of nouns", *IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, 79-84, December 2015.
5. P. Fränti, K. Waga, and C. Khurana, "Can social network be used for location-aware recommendation", *Int. Conf. on Web Information Systems & Technologies (WEBIST'15)*, 558-565, 2015.
6. N. Gali, A. Tabarcea, and P. Fränti, "Extracting representative image from web page", *Int. Conf. on Web Information Systems & Technologies (WEBIST'15)*,411-419, 2015

# Publications

## Clustering + recommendation + games

1. P. Fränti, R. Marescu-Istodor and L. Sengupta, "O-Mopsi: mobile orienteering game for sightseeing, exercising and education", *ACM Trans. on Multimedia, Computing, Communications, and Applications*, 13 (4), 56:1-25, August 2017.
2. M. Rezaei and P. Fränti, "Set matching measures for external cluster validity", *IEEE Trans. on Knowledge and Data Engineering*, 28 (8), 2173-2186, August 2016.
3. Q. Zhao, Y. Shi, Q. Liu and P. Fränti, "A grid-growing clustering algorithm for geo-spatial data", *Pattern Recognition Letters*, 53 (1), 77-84, February 2015.
4. A. Tabarcea, Z. Wan, K. Waga and P. Fränti, "O-Mopsi: mobile orienteering game using geotagged photos", *Int. Conf. on Web Information Systems & Technologies (WEBIST'13)*, Aachen, Germany, 300-303, May 2013.
5. Q. Zhao, M. Rezaei, H. Chen and P. Fränti, "Keyword clustering for automatic categorization", *IEEE Int. Conf. on Pattern Recognition (ICPR'12)*, Tsukuba City, Japan, 2845-2848, November 2012.
6. M. Rezaei and P. Fränti, "Matching similarity for keyword-based clustering", *Joint Int. Workshop on Structural, Syntactic, and Statistical Pattern Recognition (S+SSPR 2014)*, LNCS 8621, 193-202, Joensuu, Finland, August 2014.
7. K. Waga, A. Tabarcea and P. Fränti, "Recommendation of points of interest from user generated data collection", *IEEE Int. Conf. on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom'12)*, Pittsburgh, USA, 2012.

# PhD theses

1. Radu Mariescu-Istodor, "Efficient management and search of GPS routes", [PhD thesis](#), School of computing, Univ. Eastern Finland, August 2017.
2. Najlaa Gali, "Summarizing the content of web pages", [PhD thesis](#), School of computing, Univ. Eastern Finland, June 2017.
3. Mohammad Rezaei, "Clustering validation", [PhD thesis](#), School of computing, Univ. Eastern Finland, June 2016.
4. Karol Waga, "Processing, analysis and recommendation of location data", [PhD thesis](#), School of computing, Univ. Eastern Finland, June 2015.
5. Andrei Tabarcea, "Location-based web search and mobile applications", [PhD thesis](#), School of computing, Univ. Eastern Finland, 2014.
6. Minjie Chen, "Efficient processing and compression of map images and routes", [PhD thesis](#), School of computing, Univ. Eastern Finland, August 2012.
7. Qinpei Zhao, "Cluster validity in clustering methods", [PhD thesis](#), School of computing, Univ. Eastern Finland, June 2012.

# MSc Theses

1. Chaitanya Khurana, "Determining User Influence on a Social Network ",  
**MSc thesis**, UEF 2015
2. Anton Tsypchenko, "Automatic Game Generation for O-Mopsi Mobile Orienteering Game",  
**MSc thesis**, School of computing, UEF 2016.
3. Zhentian Wan, "O-Mopsi: Location-based orienteering mobile game",  
**MSc thesis**, School of computing, UEF, 2014
4. Joni Pakarinen, "Optimization of home care services",  
**MSc thesis**, School of computing, UEF, 2014.
5. Radu Mariescu-Istodor, "Detecting user actions in MOPSI",  
**MSc thesis**, School of computing, UEF, August 2013. (5/5)

**Old Topics**  
(not active at the moment)

# Augmented reality with Mopsi

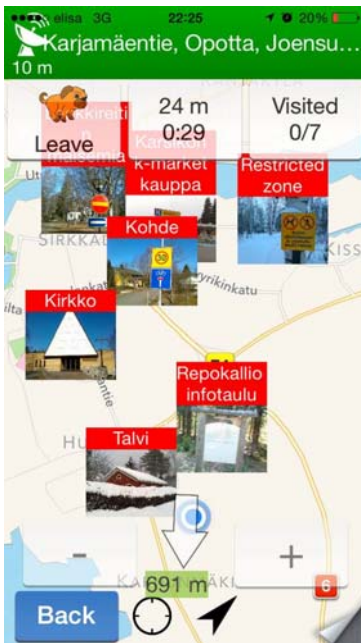
- Goal would be to integrate this to Mopsi for enhancing UI
- Game in smart phone (with GPS)
- Connect via BT
- Extra interface:
  - (1) Mopsi data from phone to device
  - (2) User input (gesture, other ways) guiding Mopsi
- Extra output on the lenses can be Mopsi services, user data from database, or other users if happen to met in real life.

**Native, Cocos, Unity?**



# Augmented reality for O-Mopsi

- Same as above but guiding the game playing
- For example showing virtual flag on visor
- Optionally to create new game but requires [http://www.vastavalo.net/albums/userpics/12793/normal\\_EK\\_1566\\_edited-1.jpg](http://www.vastavalo.net/albums/userpics/12793/normal_EK_1566_edited-1.jpg)
- Would easily create ideas for new VR/AR gaming





# Modeling of distributions based on samples

- Geo-tagged data collected from field
  - Hundreds of samples at each location (GPS)
  - Raw data is spectral measurement of reflections
  - Processed data is intensity of detected elements (alkuaine)
- Device connected to phone (BT) and server (network)
- Interactive data handling on map in server
  - Zooming & panning
  - Choosing elements
  - Modeling distributions based on samples

