Thesis topics

Pasi Fränti

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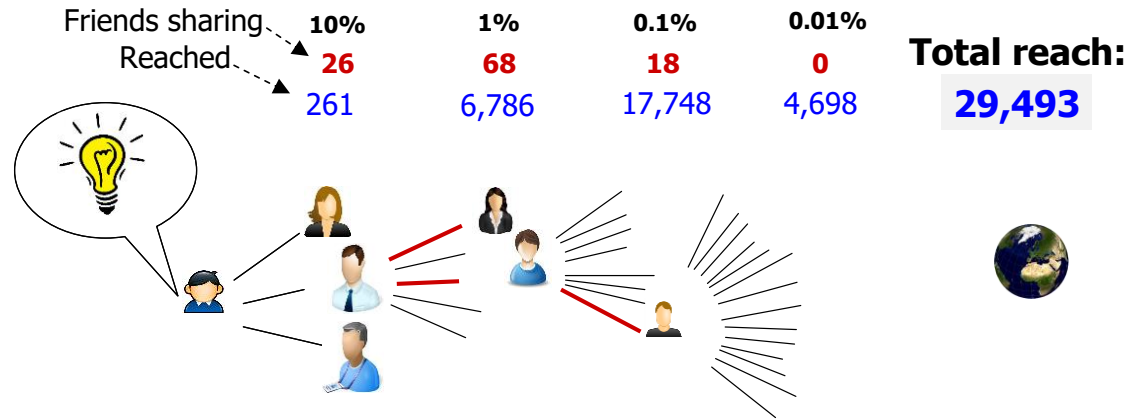University of Eastern Finland
Joensuu, FINLAND

Past students

2020

- Rushikesh Sane: Density-based clustering V.2 (**completed**)
- Raheel Mansoor: Route prediction (**completed**)
- Liu Yunlong: Word cloud on map (**completed**)
- Tania Akter: Searching Mopsi data (**completed**)

2021

- Hiep Nguyen: Clustering trajectories (**completed**)
- Awais Akram: Clustering STEMI patients in IMPRO (**completed**)
- Ashfaq Afzal: Visualizing Clustering Algorithms (**completed**)
- Shariful Majumar: Multi-lingual keyword extraction (**completed**)
- Abedin Bhuiyan: Collecting ground truth for music mood (**completed**)
- Shenal Jayratne: Automatic creation of home page (**completed**)
- Imranul Ismal: Image extraction from web page (**completed**)
- Erfan Ahmed: POS patterns in Tweets (**completed**)

Web mining

Ad hoc and social networks analysis

	10%	1%	0.1%	0.01%	
Friends sharing	**26**	**68**	**18**	**0**	**Total reach:**
Reached	261	6,786	17,748	4,698	**29,493**

1. Ah 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: Sijantitiedon 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.
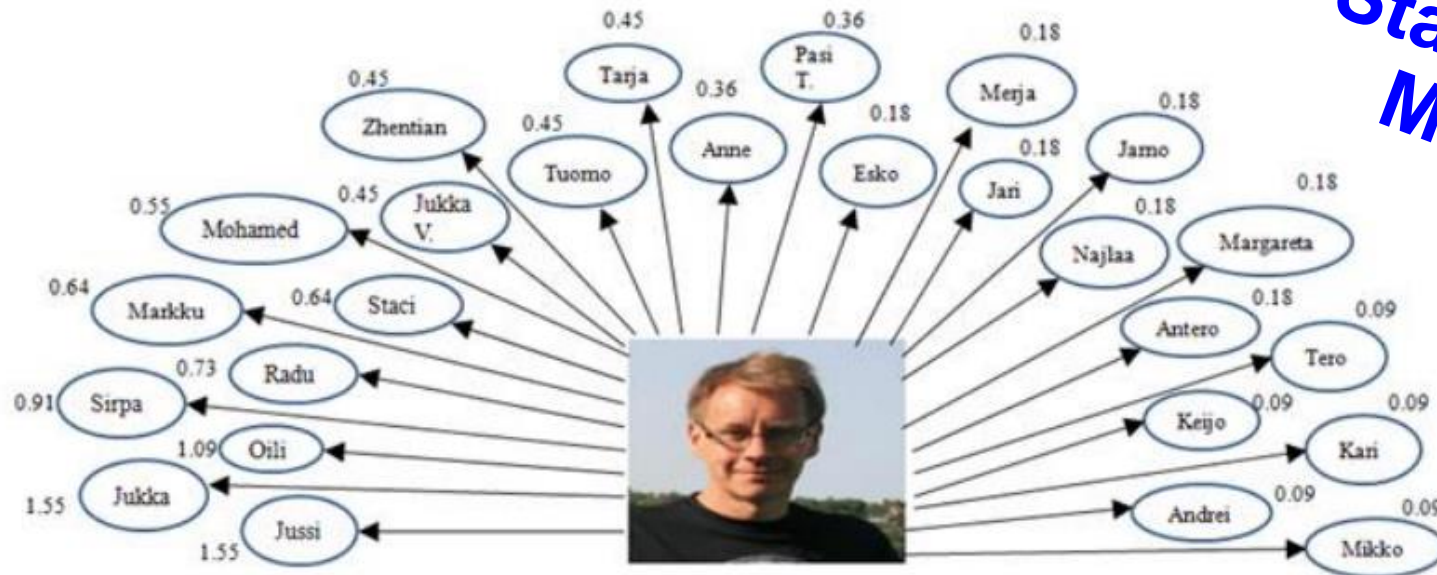
Pasi's influence = **11.9**

Language-independent content extraction

Logo image →

Navigation bar

Title →

keywords:
Started
SM

Keywords

Text →

NLP tools for detecting tweeting Bots
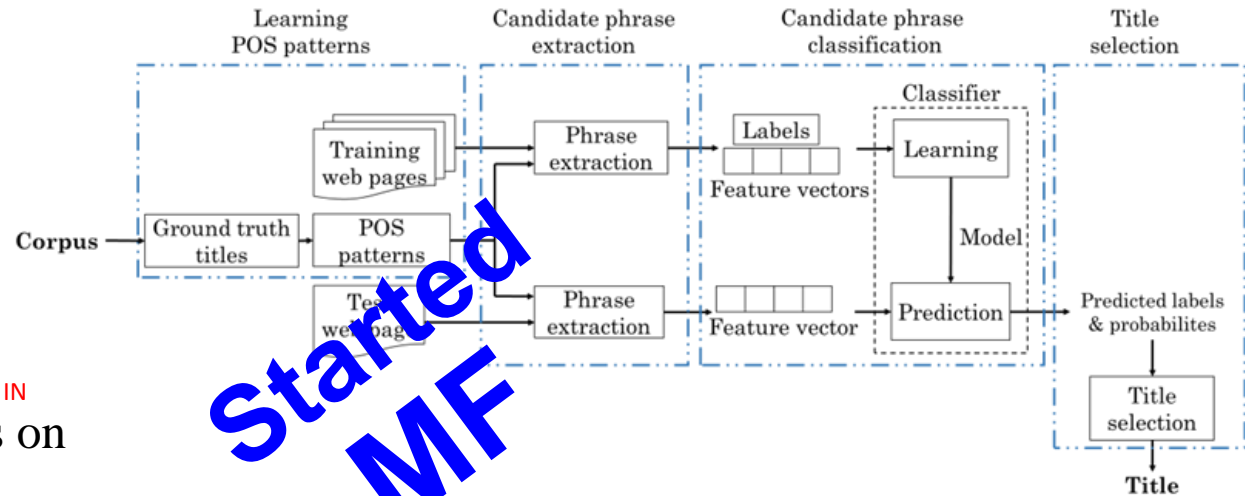
Project on-going
http://cs.uef.fi/sipu/OpenPositions-2.pdf

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-occurences of the words. Estimate similarity based on this and their total frequence w/o any language information.

Example: **Uutinen**, **news**
High similarity because same meaning despite different language.
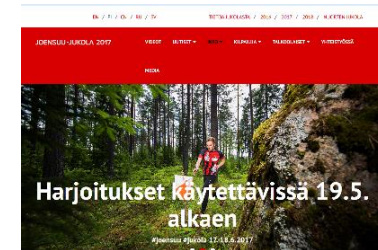
# Detecting clusters from text phrases

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
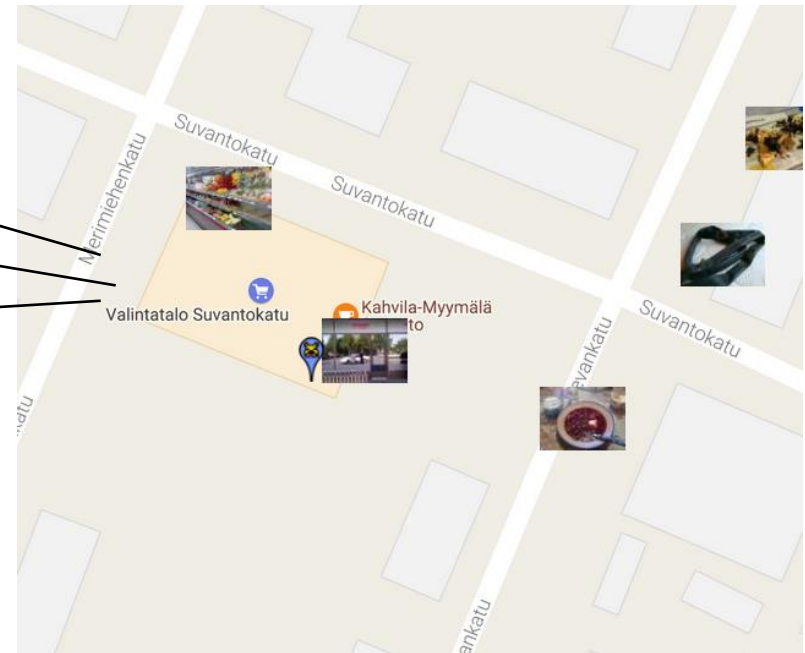
Clustering

Split k-means

- Split/Divisive algorithms build the clustering hierarchically by dividing clusters one by one until k clusters found.

- Split + K-means apply k-means after every split. Locally for the new clusters, or globally.

- Three design choices: (1) what cluster to split next; (2) how to split; (3) tuning the result.

- Study the variants what matters and what not. Is it enough to do random split and k-means, or is more sophisticated splitting needed?

- Earlier results (Fränti&Kaukoranta) showed that choice of the components was important while k-means was considered merely optional. Simple variants with k-means not studied.

- Hypothesis is that the split maybe not be important if we use k-means every step. But then efficiency of k-means is important, otherwise the speed benefit would be lost.

- When solving clustering for several k-values, which is better:

  - Apply Split algorithm for all k=1..N

  - Apply Random swap for all k=1..N

See also clustering course material: *Divisive algorithms* slides.
http://cs.uef.fi/pages/franti/cluster/

Started: OR

P. Fränti, T. Kaukoranta and O. Nevalainen: "On the Splitting Method for VQ Codebook Generation", *Optical Engineering*, 1997.

M. Steinbach, G. Karypis, V. Kumar, A comparison of document clustering techniques, in KDD workshop on text mining, 2000.

S-S. Yu, S-W. Chu, C-M. Wang, Y-K. Chan, T-C. Chang, Two improved k-means algorithms, *Applied Soft Computing*, 2018.

Q. Zhao and P. Fränti, "WB-index: a sum-of-squares based index for cluster validity", *Data & Knowledge Engineering*, 2014.

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)
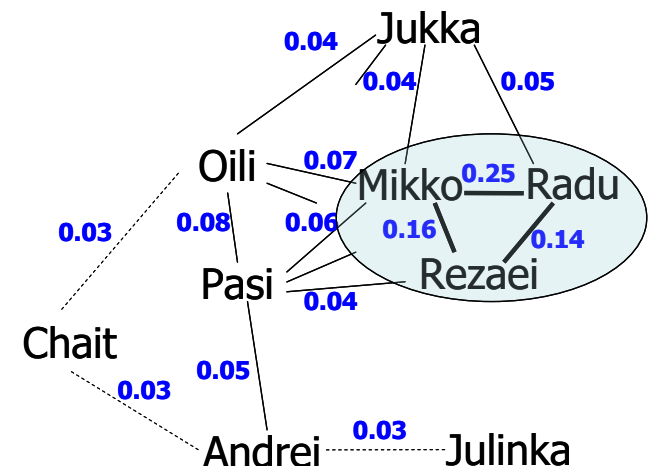- Based on similar interests (likes in FB)

Recommendation:

- Relevance = Influence of user * Relevance of POI
- Influence = Activity + dominance

Approach A:

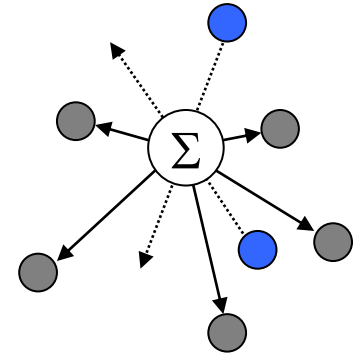- Group users together
- Same recommendation for all users

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
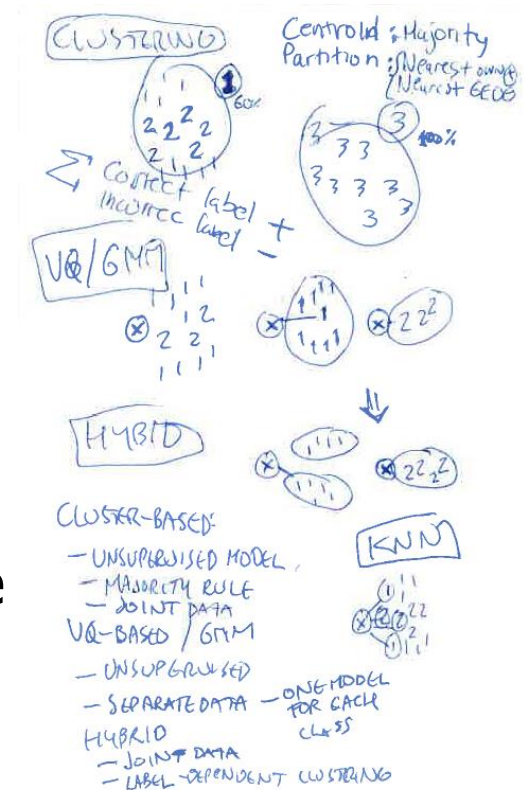
Clustering categorical data

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

Entropy-based variant: MA

- 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: $SUM(1/d) = max!$
  - $MSE = SUM\_ij(d^2)/n = min!$
  - $MSE = SUM\_ij(1/d^2)/n = max!$

TSP

TSP algorithms
Suitable also for DAA projects

1. Christofides   **Completed**

2. Double-tree

3. Kruskal TSP   **Completed AK + HN**

• Similar to Kruskal but adding links only to the end of chains

4. MST → TSP   **Completed TN**

• Detect and break branches and re-connect via leaves

5. Local search using spanning trees

• Use 2-opt, node swap, link swap operations

• Allow branches in the intermediate solutions

6. Local search using node clusters

• Operates on clusters instead of single nodes.

7. Relaxed variant of TSP

• Normal variant enters/exits node using one link exactly.

• Relaxed variant allows fractional enter/leaving (25%, 35%, 40%)

Stopping criterion for local search

- Random Mix local-search shows good quality up to N=30 targets with the following parameters:
  - 25 repeats
  - 10 000 iterations
- Investigate how many repeats and iterations are required for an instance with a given number of nodes. Requires knowledge of Java (or C).
- Invent stopping criterion based on factors like:
  - Problem size (N)
  - Complexity of the problem
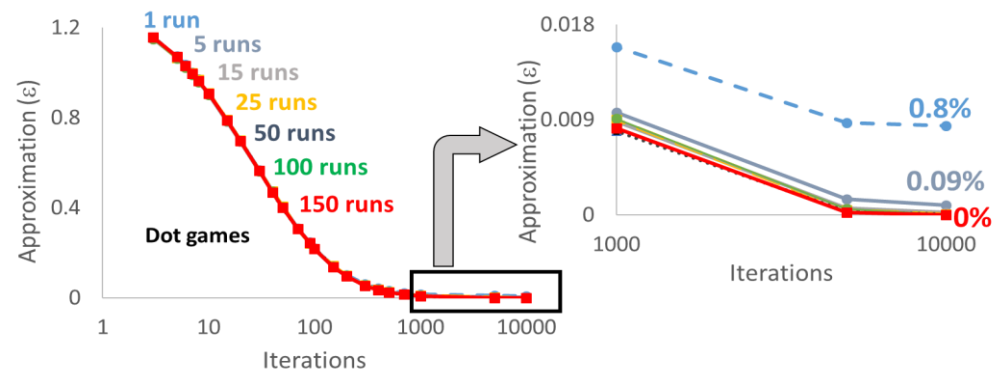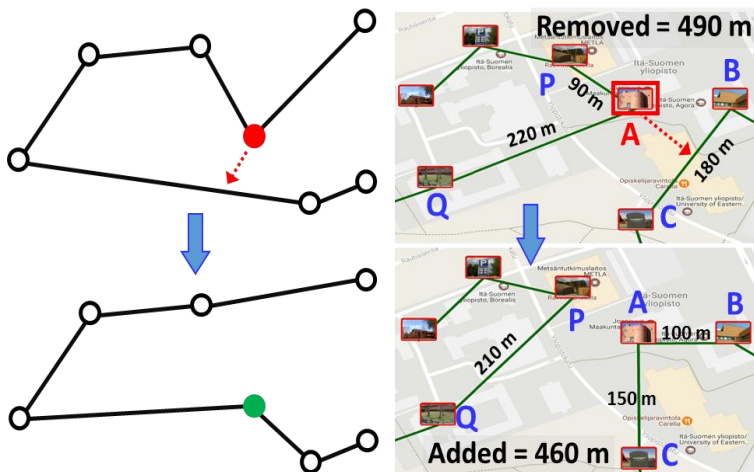
L. Sengupta, R. Mariescu-Istodor and P. Fränti
Which local search operator works best for open loop Euclidean TSP
Applied Sciences, 9 (19), 3985, 2019.

Human-aided TSP solver

Build an interface where user can interact with a TSP solver by telling it where to find improvements. To solve the TSP locally, apply an algorithm like [1].

User should be able to interact by:

- Clicking somewhere in the space

- Selecting a subset of targets (lasso tool)

- Have multiple regions and solve in parallel

Started: JT
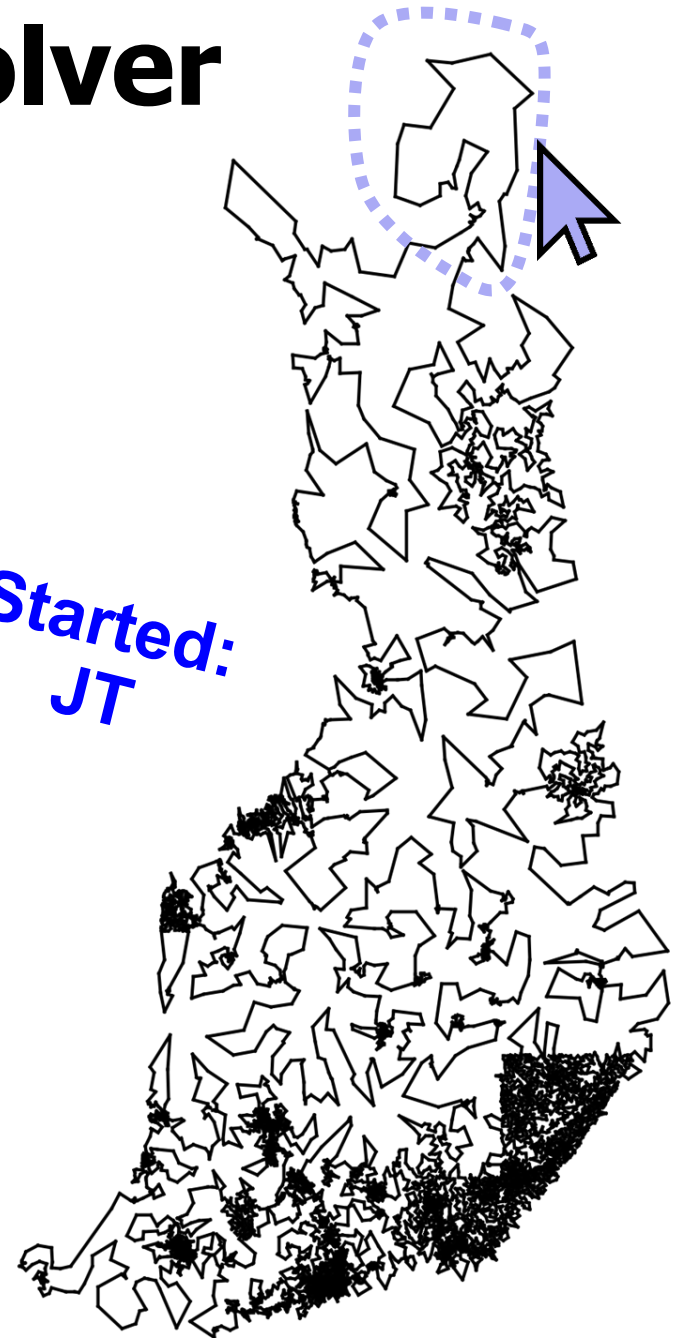
[1] Sengupta, L., Mariescu-Istodor, R., & Fränti, P. (2019). Which Local Search Operator Works Best for the Open-Loop TSP?. *Applied Sciences*, 9(19), 3985.

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
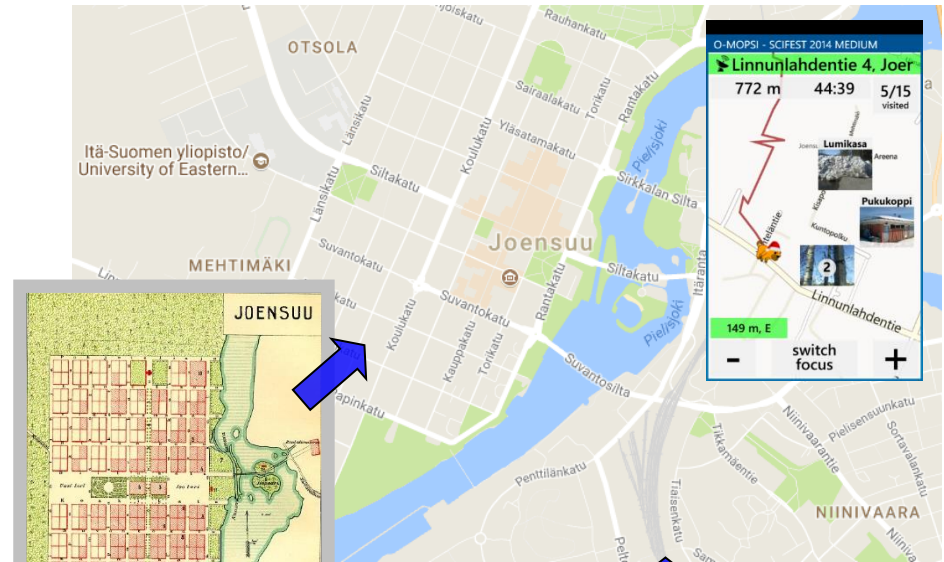- Ready to learn Android
- Might require web-programming
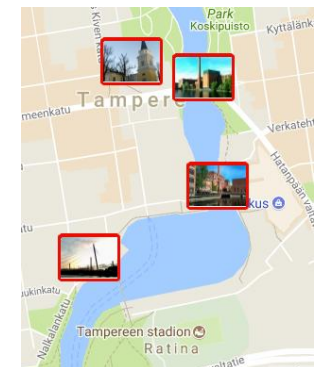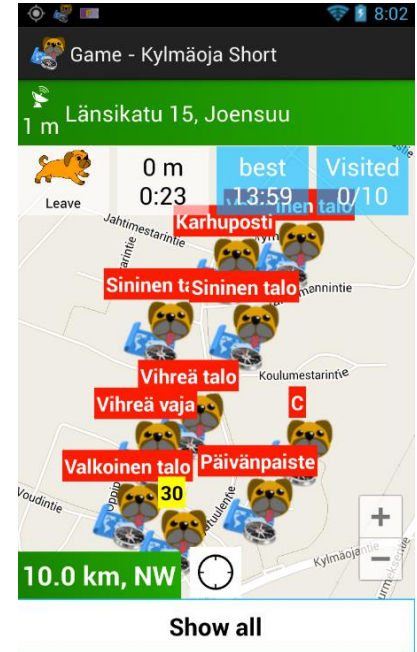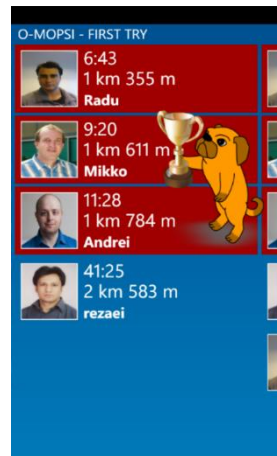
Educational applications

Tools for adding material:

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

Educational games?

Automatic content creation

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

Outdoor museums

Geography

Botanical gardens

Biology

Architecture

City history

Exhibitions

Aspen

Maple

Lime tree

Oak

Elm

Thuja

Spruce

Birch

Rowan

Automatic content creation

- Expanding O-Mopsi database by geo-tagged photos.

- Improve the user interface by adding ways for the users to interact with this collection to facilitate game creation (selection of targets).

- Invent tools for analyzing picture quality.

- Required: PhP, JavaScript and MySql.

- Good to know: HTML, CSS and Python.

Started: MII + LK

Lifesaver

Museum

Swimmer boys
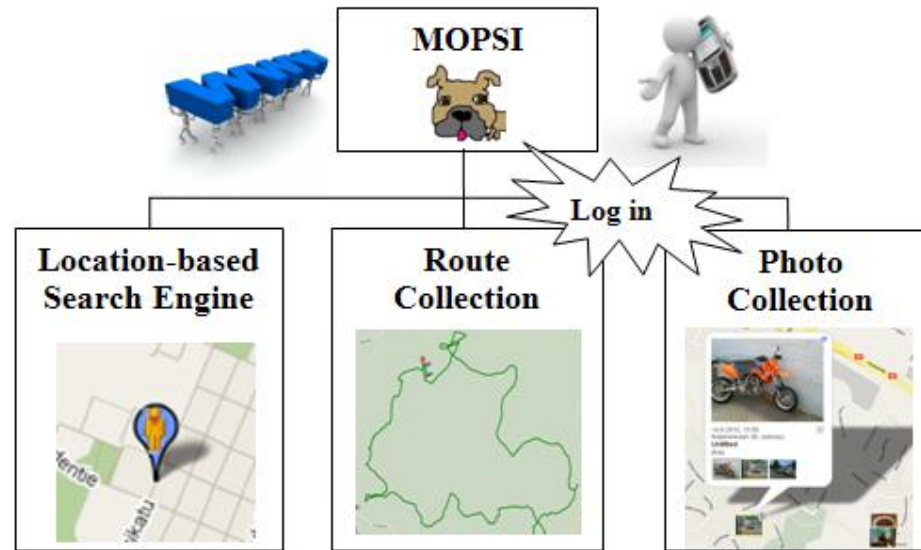

Graduation stone

Start of route

Freedom bridge

Science Park → Science Park
Maybe / Better

Church → Sakasti
Reconsider / Better

Mopsi

Mopsi overview

MOPSI

Log in

Location-based Search Engine

Route Collection

Photo Collection

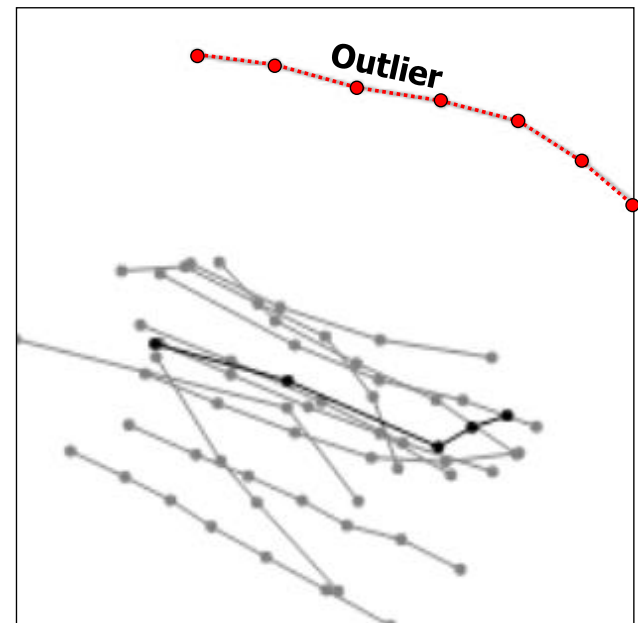| Recommendation System	Web Content Mining	Route Pattern	Social network	Location-based Game
➢ Service (bus, friend) ➢ Text (search query, photo description)	➢ Meta searching ➢ Service title detection ➢ Document processing	➢ Route reduction ➢ Route segmentation ➢ Activity area	➢ Facebook	➢ Orienteering ➢ Killer-game

Detecting outliers among GPS trajectories

- Apply existing k-NN based methods: (KNN, ODIN, MOD, LOF)
- Generalize from points to GPS trajectories
- Compare different distance functions
- Evaluation:
  - Does it improve quality of segment averages
  - Does it improve quality of clustering
- To get started:

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

- Requires knowledge of:
  - Data structures
  - Algorithms
  - Python (Java)

Started: BJ

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.

Mopsi database reporting tool
SQL-oriented topic

To provide insights of what is going on like number of daily/monthly users, and to make reports of queries like:

1. Find Pasi's photos having description "Lenkkireitin maisemia".

2. Find all text descriptions that both Pasi and Radu has used.

3. Find Pasi's most popular text descriptions.

4. Find photos with the "Lenkkireitin maisemia" and then find if any other user has taken photos at the **same time and place** as Pasi.

Evaluating user experience
# In location-based services

- Aspects like design, performance, usability and location

- Two approaches can work:

    1. Crowdsourcing

    2. Collecting user feedback data from existing services using APIs (or if dataset available)

- Literature-oriented.

Started: JH

Clustering routes

Medoid-based: completed

Route prediction
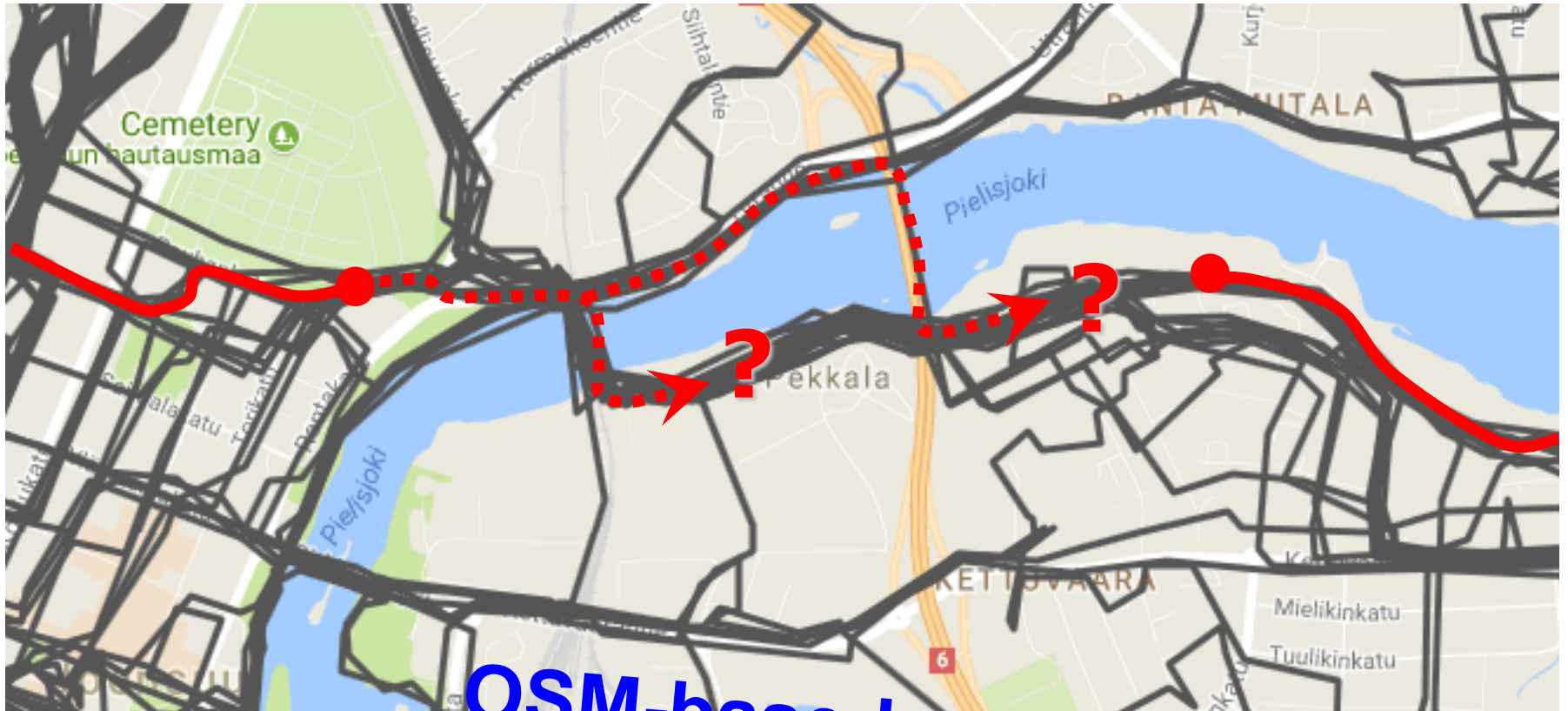
OSM-based: completed
AI-approach: open

Organizing route collection for efficient retrieval

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+3-years on-going research project

Started: JJ+AA
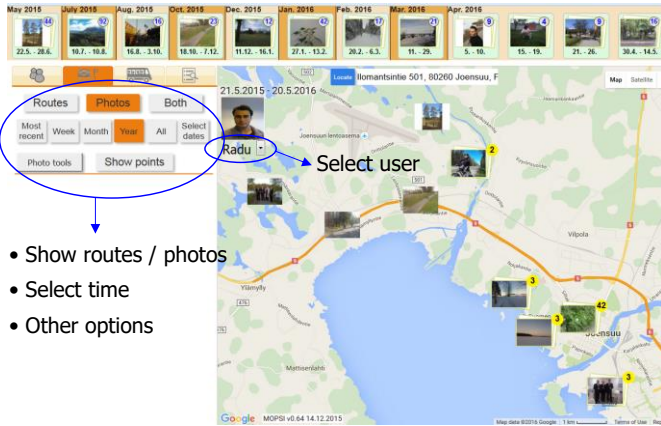
- Show routes / photos
- Select time
- Other options

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

Location-aware recommendation

Given location, recommend Relevant places around

Vilkku kahvio

Location	0.55
Time	1.00
Relevance	0.84
User network	0.57 ⭐⭐⭐
Total:	**2.95**

Keywords	All		
	Nearby	Recently	Total
kahvila	0.32	0.00	0.27
Total (Max)	0.32	0.00	0.27
Total (Av)	0.32	0.00	0.27

Kuurnankulma

Location	0.60
Time	1.00
Relevance	0.32
User network	1.00 ⭐⭐⭐⭐⭐
Total:	**2.92**

Keywords	All		
	Nearby	Recently	Total
lounas	0.02	0.00	0.02
Total (Max)	0.02	0.00	0.02
Total (Av)	0.02	0.00	0.02

Vilkku kahvio

740 m

306 m

Kuurnankulma

762 m

Heinosen leipomo

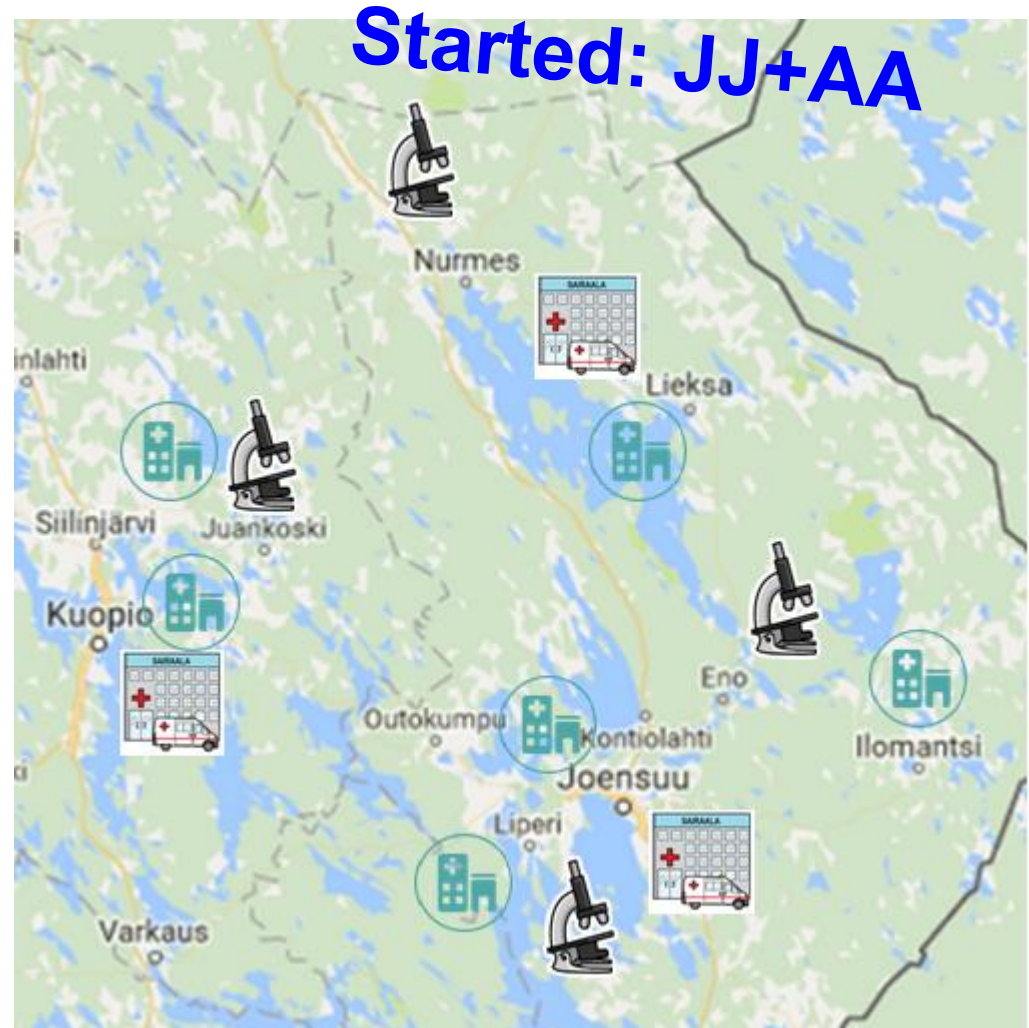|	
Location	0.95
Time	1.00
Relevance	0.84
User network	0.00
Total:	**2.79**

Keywords	All		
	Nearby	Recently	Total
kahvila	0.32	0.00	0.27
Total (Max)	0.32	0.00	0.27
Total (Av)	0.32	0.00	0.27

Vilkku Kahvio
Pamilonkatu 33, Joensuu
740 m
Recom info
⭐⭐⭐
Check route

Kuurnankulma
Paukkajantie 2-4, Joensuu
762 m
Recom info
⭐⭐⭐⭐
Check route

Kahvila Heinosen leipomo
Hiiskoskentie 13, Joensuu
306 m
Recom info
⭐⭐⭐⭐⭐
Check route

La Dolce Vita
Kuurnankatu 6, Joensuu
625 m
Recom info
⭐⭐⭐⭐⭐
Check route

Kahvila Huili & Javerstok grilli
Kuurnankatu 14, Joensuu
420 m
Recom info
⭐⭐⭐
Check route

Skarppi - ylioppilaskunnan sauna
Kaislakatu 10, Joensuu
758 m
Recom info
⭐⭐⭐

* Search keywords used (total)
* Search keywords used (recently)
* All service ratings given
* Likes of other user photos
* Keywords in the Services created
* Words used in photo description

Google MOPSI v0.63 24.3.2015

Other topics

Detecting features from song

- Extract features from the song
  - Rhythm
  - Timbre
  - ...
- Use existing tools as much as possible
- Possible extension: apply speech-to-text and then sentiment analysis on text via natural language processing.

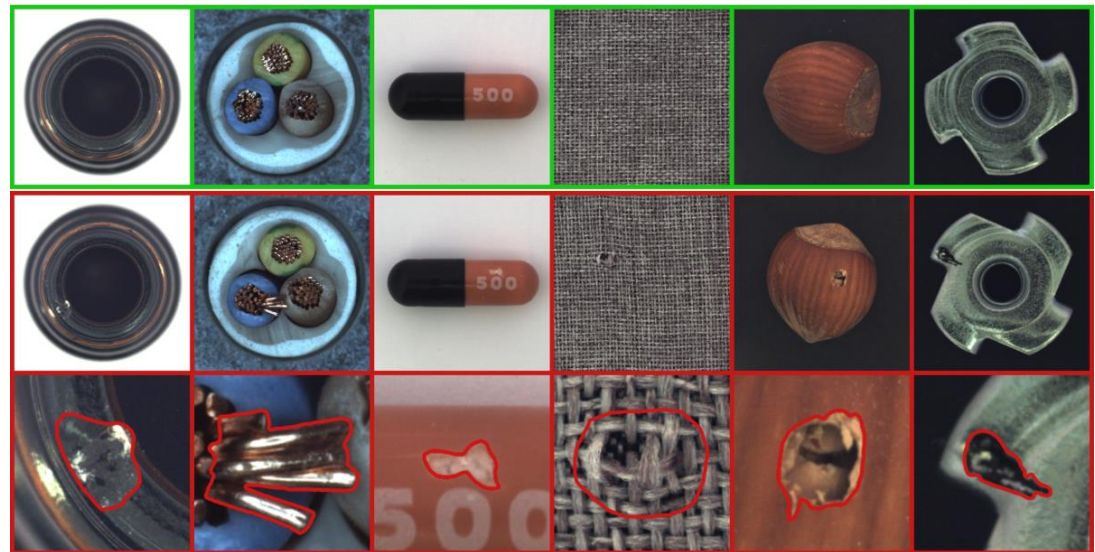Melody is on "strong" beats - not Syncopated
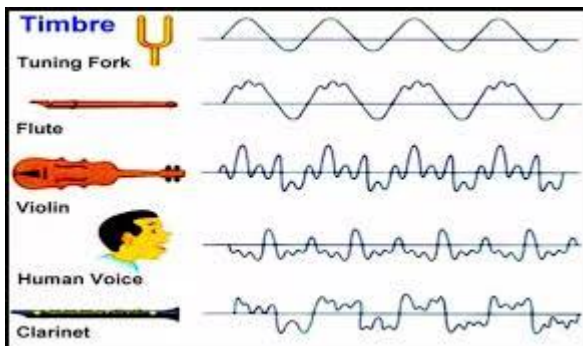
Anomaly detection of image patches

1. Apply k-NN based outlier detection to image patches
2. Study the effects of:
   - Detector
   - Neighborhood averaging
   - Choice of threshold
3. Implementations exist
4. Collaborating Giacomo Univ. of Milan, Italy
5. Outlier detection literature:

Carrera D., Manganini F., Boracchi G., Lanzarone E. "Defect Detection in SEM Images of Nanofibrous Materials", *IEEE Transactions on Industrial Informatics,* 2017,

J.W. Yang. S. Rahardja and P. Fränti, "Mean-shift outlier detection", FSDM, 2018

J.W. Yang. S. Rahardja and P. Fränti, "Outlier detection: how to threshold outlier scores?", AIIPCC, 2019.

https://www.mvtec.com/company/research/datasets/mvtec-ad/

Anomaly detection (continued)

Monitoring Nanofiber Production

Carrera D., Manganini F., Boracchi G., Lanzarone E. *"Defect Detection in SEM Images of Nanofibrous Materials"*, IEEE Transactions on Industrial Informatics 2017, 11 pages, doi:10.1109/TII.2016.2641472
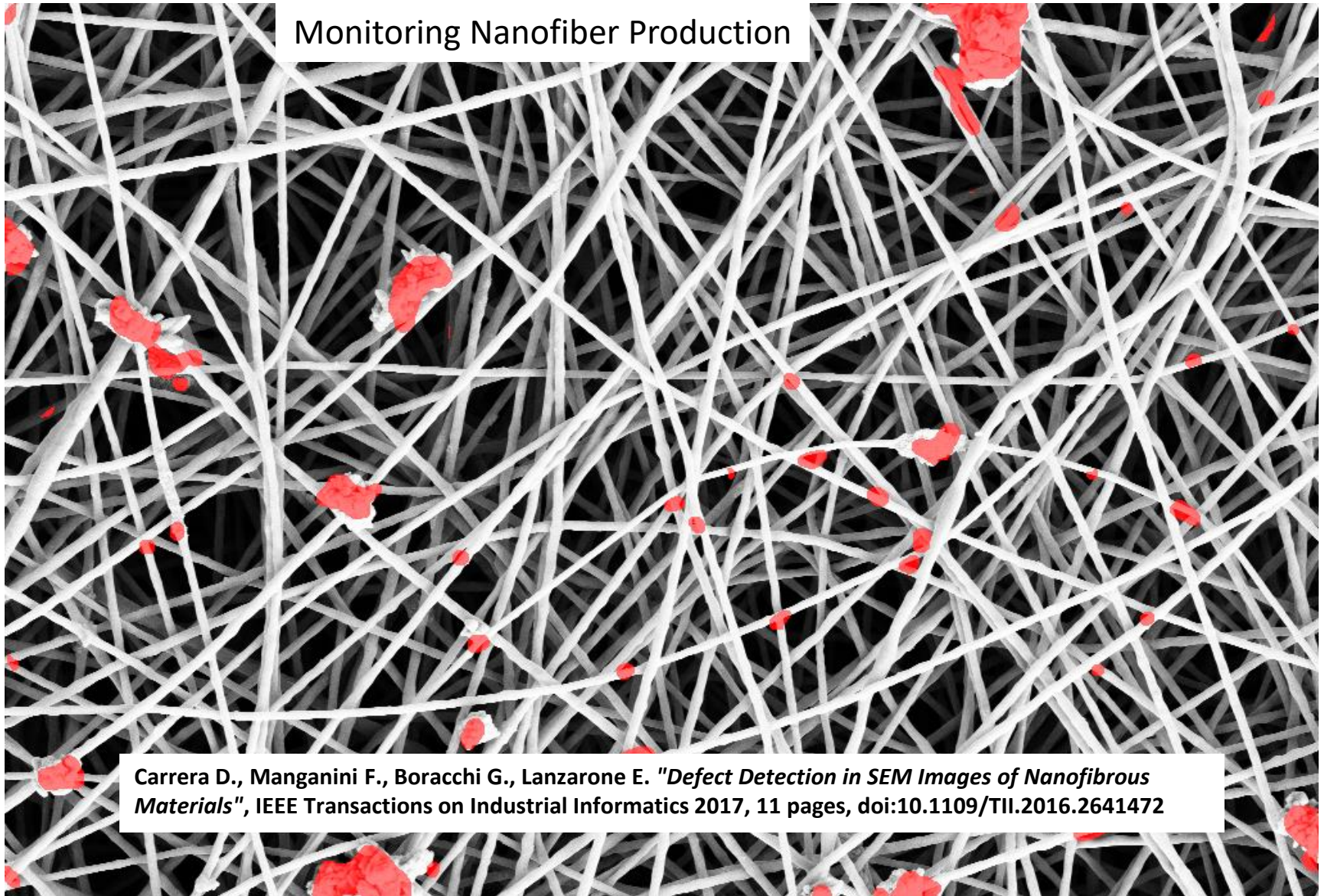
Completed Topics

(just those topics originally announced here)

Labelling music in databases

- Web application that can load and play music.
  - Next song
  - Previous song
  - Play, Pause, Skip
- Tool to label using pre-defined categories.
  - Emotion, genre, …
- Results must be stored on a server.
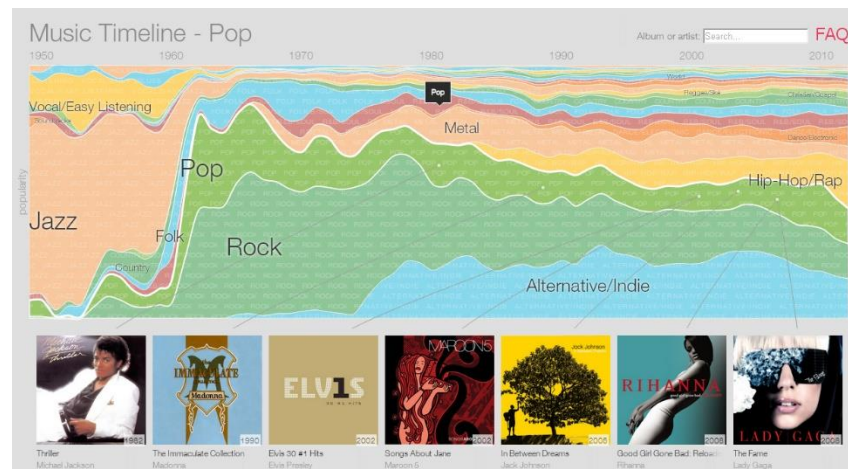- Requires: PhP and JavaScript.
- Good to know: HTML, CSS and MySQL.

Completed

RHYTHM AND VOICE

...for the soul

Automatic creation of home 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/

Completed SJ

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
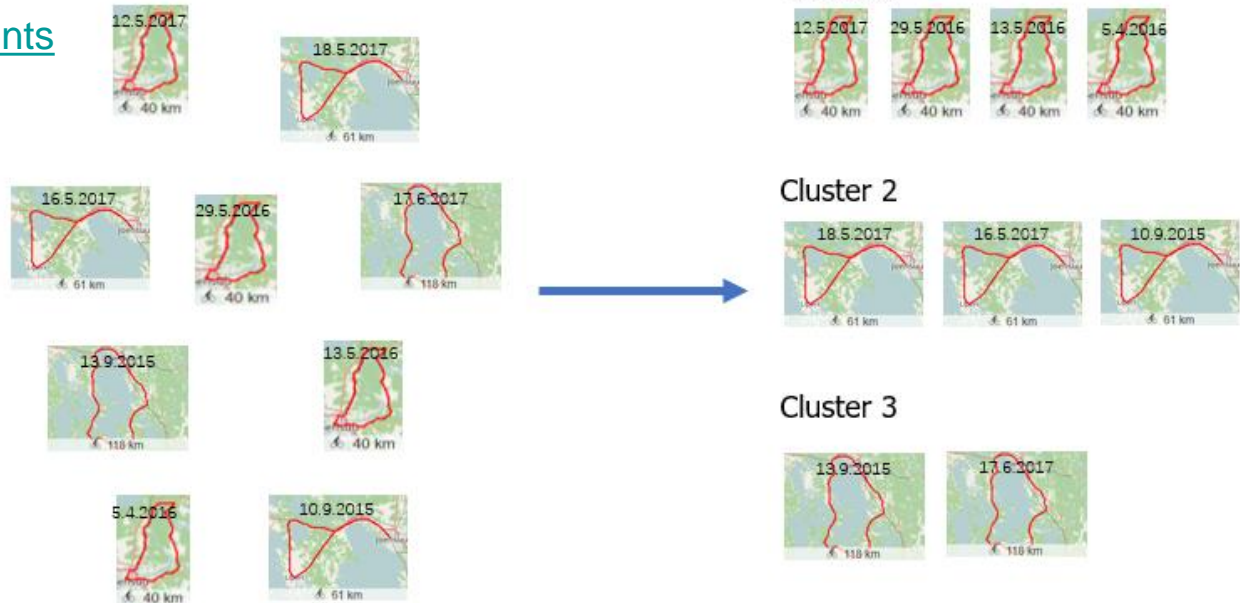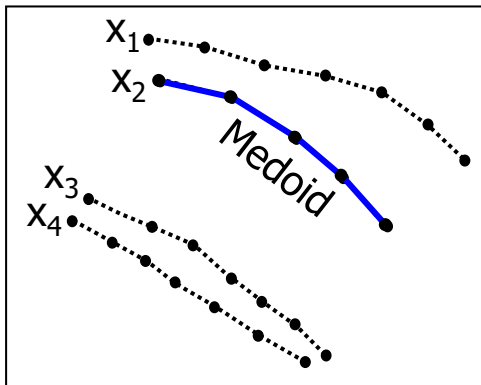- Best
- Pizza

Niskakatu 2, 80100 Joensuu

K-medoids in segment averaging and clustering

- Systematic study how well medoids represent segment average.
- Effect of parameters:
  - (a) size of set
  - (b) choice of distance function,
  - (c) presence of outliers.
- Two applications:
  - 1) segment averaging;
  - 2) clustering.
- http://cs.uef.fi/sipu/segments
- See notes below!

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
- λ is increasing until converged to balance solution

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

Completed

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)
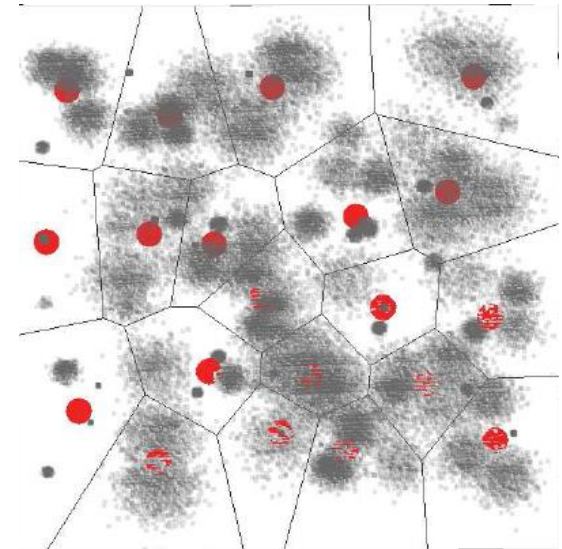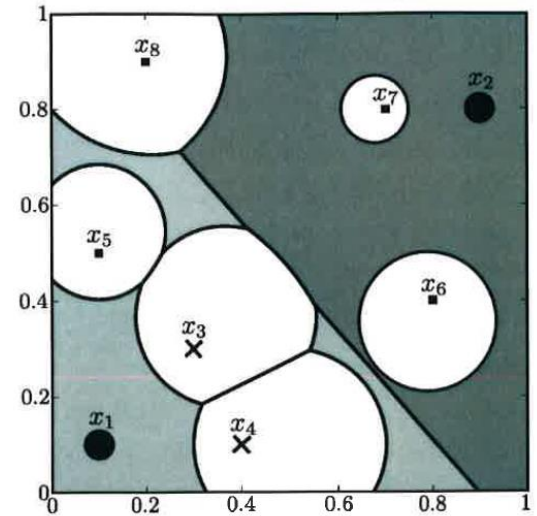- Resource allocation problem

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

Mobile bus time table systems

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
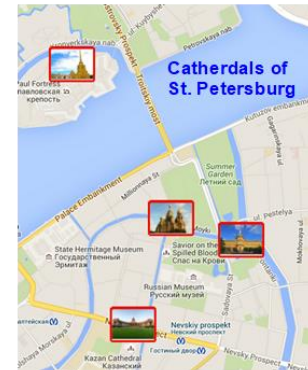
Completed

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

Lenkille.com

Kimppalenkkien (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.
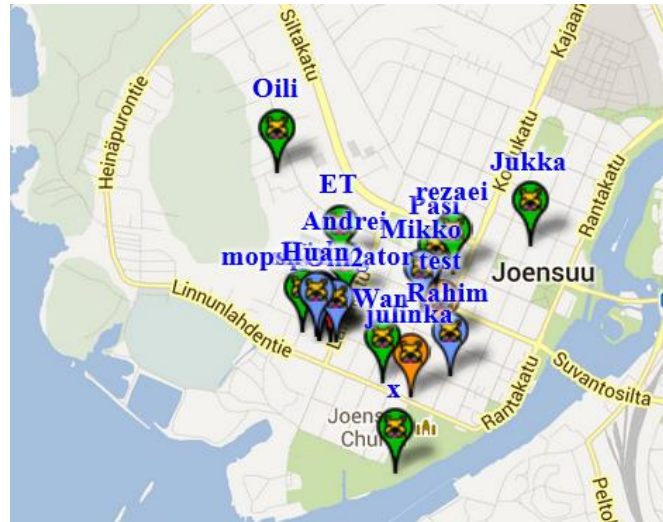
Planned events:

- 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

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

Completed
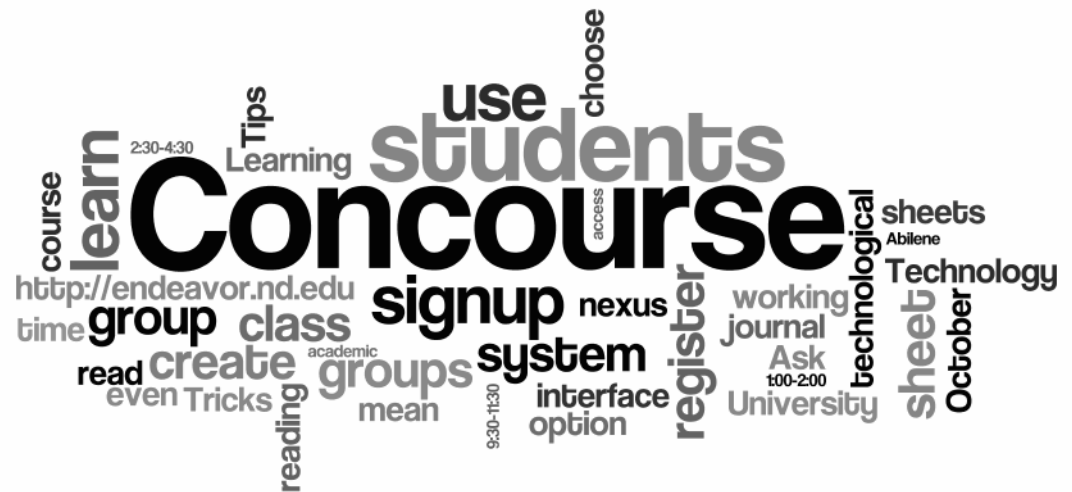
Old Topics
(not active at the moment)

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
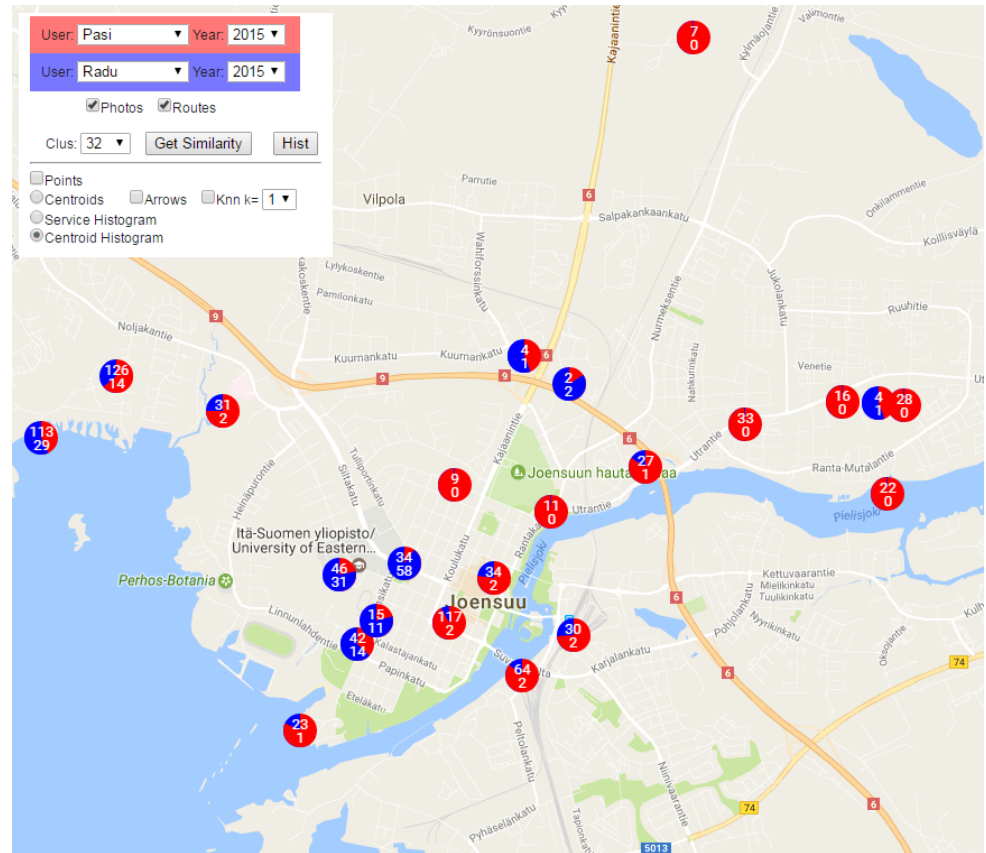
Augmented reality with Mopsi

- 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?
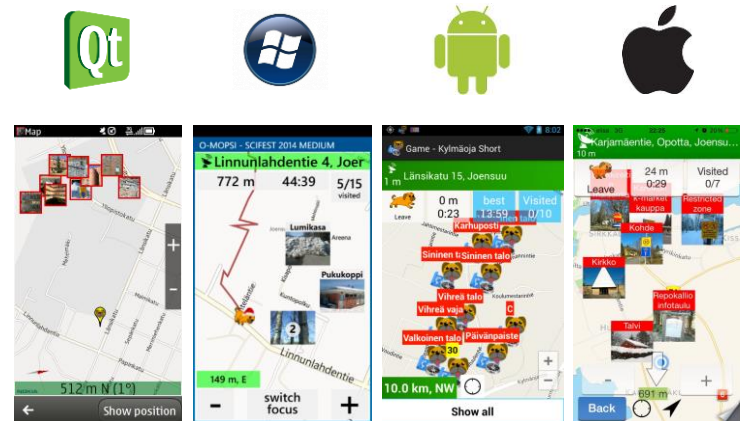
Augmented reality for O-Mopsi

- 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
- Would easily create ideas for new VR/AR gaming
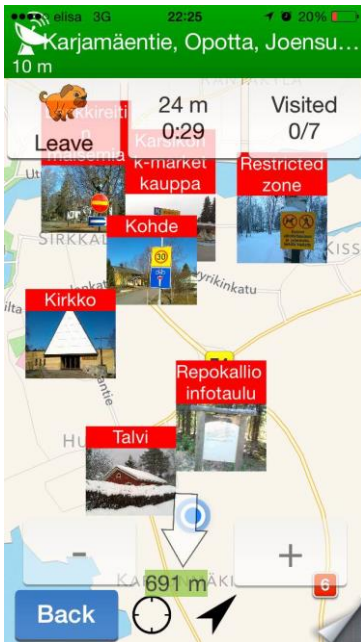
Modeling of distributions based on samples

- 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

Conventional method · Our method

30 km

30 km

Deposit
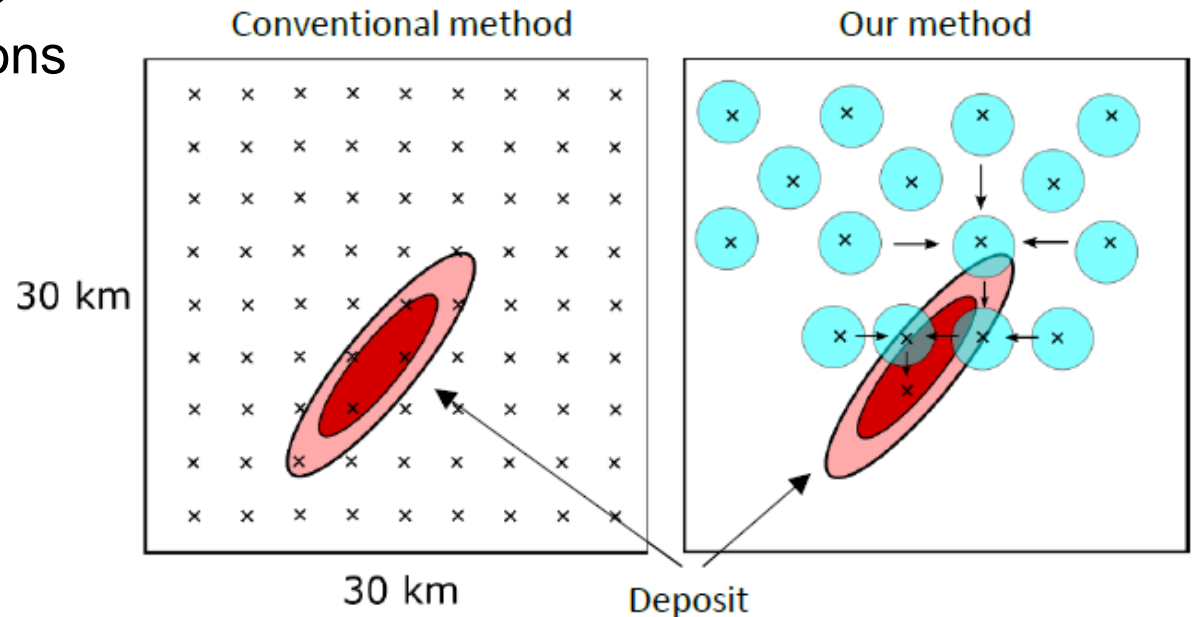
Publications
Location-based related

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. Mariescu-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. Mariescu-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, "ClRank: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. Mariescu-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, "Summarinzing 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)