



Christmas seminar 2018

Social and health care services as an optimization problem

Pasi Fränti

Univ. of Eastern Finland
Joensuu, FINLAND



UNIVERSITY OF
EASTERN FINLAND



MACHINE
LEARNING



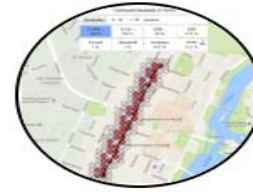
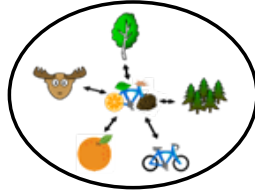
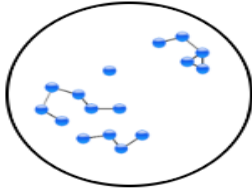
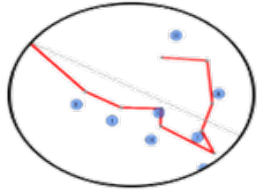


Data mining
Information retrieval
Location-aware applications



Location-aware applications

Clustering

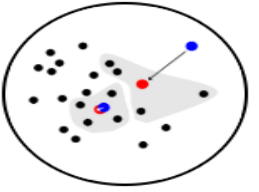
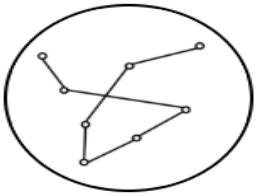


GPS trajectories



Spatial data

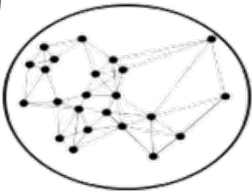
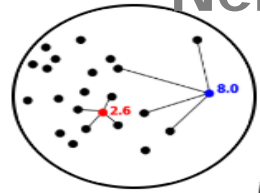
TSP



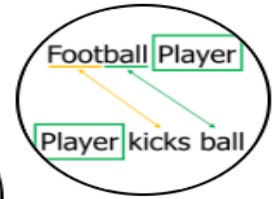
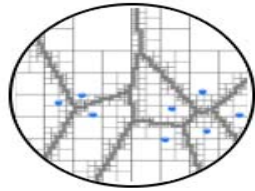
Games



Neighborhood graphs



Clustering



Text mining



Move type detection

K. Waga, A. Tabarcea, M. Chen and P. Fränti,
“Detecting movement type by route segmentation and
Classification”, *CollaborateCom*, Pittsburgh, USA, 2012

Movement types considered:



Walk



Run



Bicycle



Car

Other possibilities:



Train



Boat



Flight



Skiing



Bus

Spatial context
needed

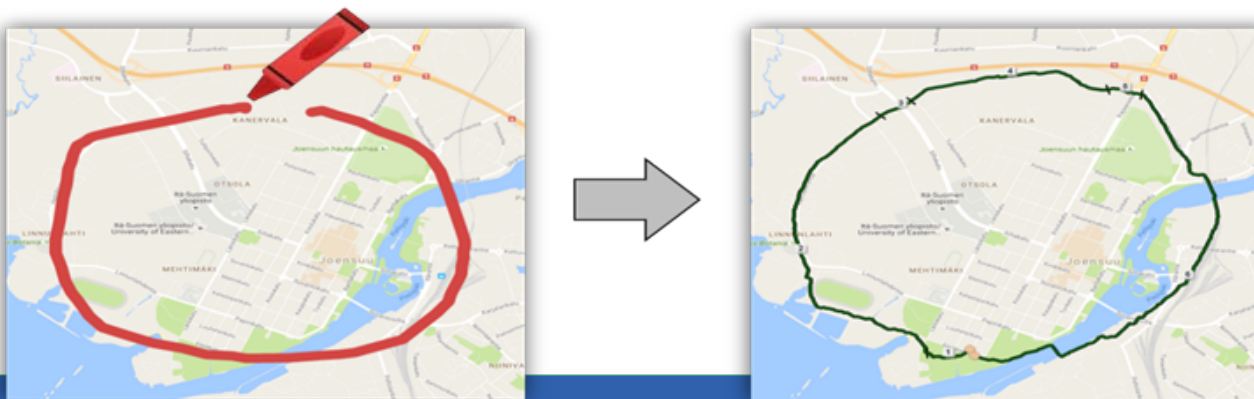
Speed?

Track location, season

Time tables

Shape search

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



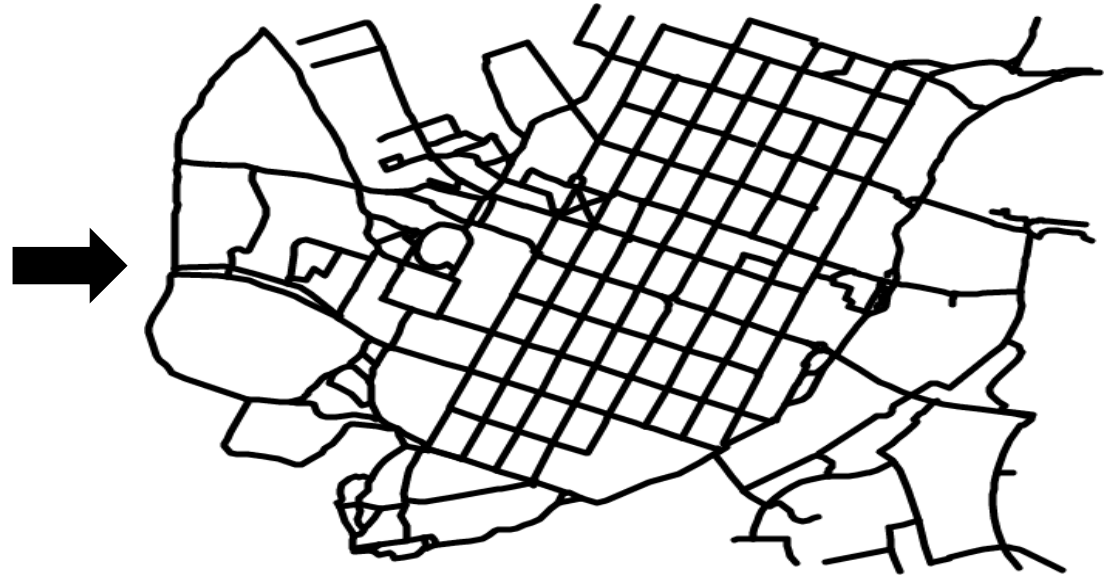
Road network extraction

R. Mariescu-Istodor and P. Fränti, "From Routes to Roads: Infer road network using GPS trajectories", *ACM Trans. on Spatial Algorithms and Systems*, 2018

GPS Trajectories



Road Network



Applications: share riding



MOPSI Events

All Users ▾

All Events ▾

Create Event

A map of Joensuu, Finland, showing a route for a running event. The route is marked with a red line and a red running icon. The route starts near the Botanical Garden, goes through the city center, and ends near the Hypermarket Prisma Joensuu. Other landmarks include the Cemetery, Ita-Suomen yliopisto, and various streets like Keskuskatu and Keskustie. The map also shows a car icon, a bicycle icon, and a bus icon. The Google logo is in the bottom left corner.

A photograph of a person running on a red track. The person is wearing a red t-shirt, grey shorts, and black running shoes. They are holding a smartphone in their right hand. The background shows green trees and a clear sky.

Map data ©2018 Google

500 m

Terms of Use

Report a map error



Web content mining

Logo image



Navigation bar

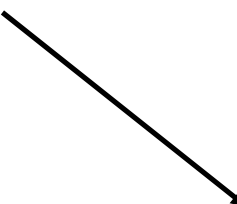
Title



Raspberry Pi 3 adds wi-fi and Bluetooth

2 hours ago | Technology

Keywords



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.

Text



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



Images



Keyword extraction

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 Web Intelligence and Intelligent Agent Technology* Singapore, 2015

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
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

Address detection

A. Tabarcea, V. Hautamäki, P. Fränti, "Ad-hoc georeferencing of web-pages using street-name prefix trees", *Int. Conf. on Web Information Systems & Technologies (WEBIST'10)*, Valencia, Spain, vol.1, 237-244, April 2010.

Pizza MASTER

★ ETUSIVU ★ JOENSUU ★ LIEKSA ★ KONTIOLAHTI
★ OTA YHTEYTTÄ

JOENSUU

LIEKSA

KONTIOLAHTI

Yhteystiedot

Niskakatu 11
80100 Joensuu
Puh. 0400 281700

KOTIIN-KULJETUS!

ma-to 10:30-22:00
pe-la 10:30-04:30
su 12:00-22:00

PIZZA MASTER JOENSUU

Yhteystiedot

Pielisentie 10-12
81700 Lieksa
Puh. 0400 288481

MAUKAS LOUNAS!

ma-to 11:00-21:00
pe-la 10:30-05:00
su 12:00-21:00

PIZZA MASTER LIEKSA

Yhteystiedot

Keskuskatu 21
81100 Kontiolahti
Puh. 0400 732030

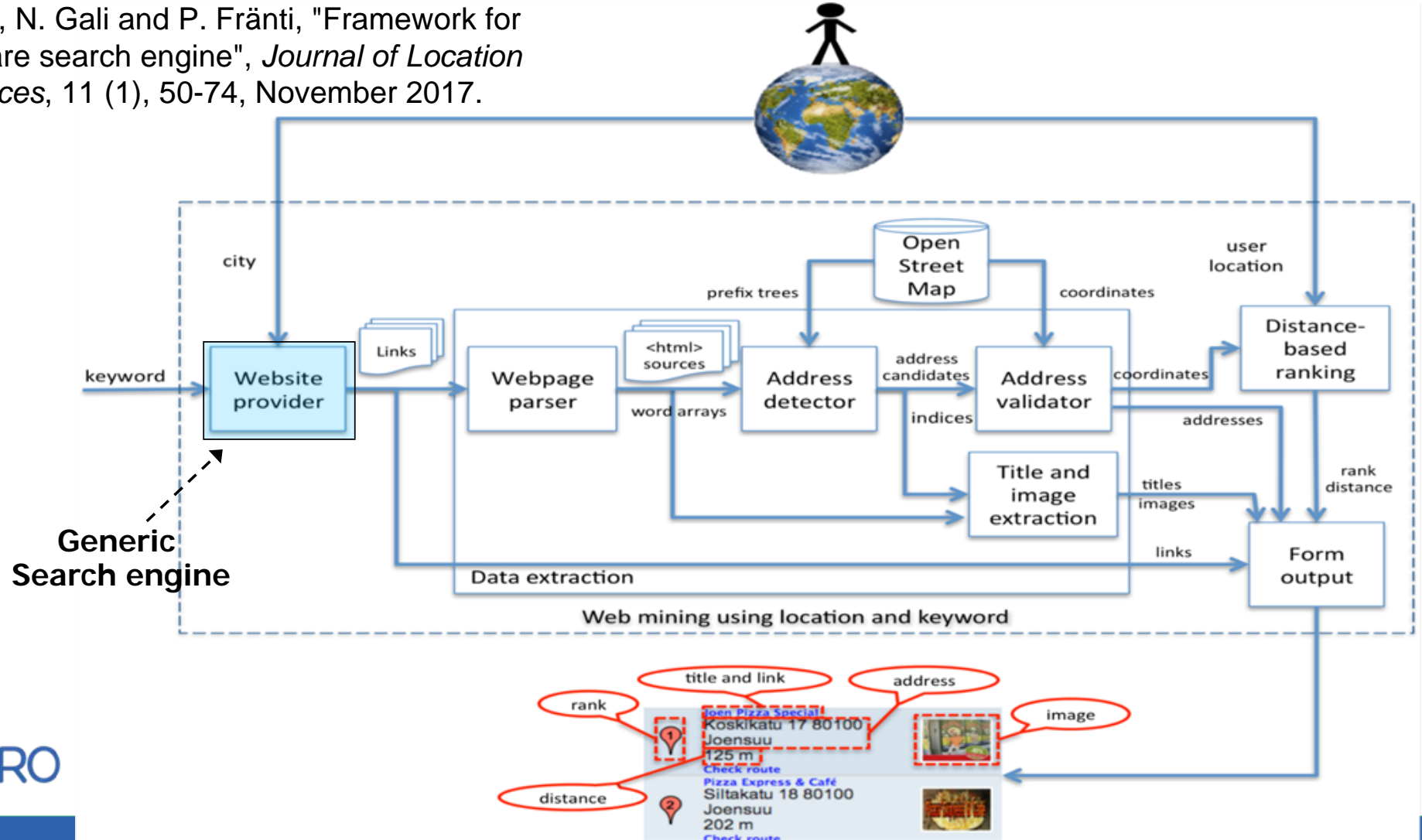
ma-la 11:00-21:00
su 12:00-21:00

PIZZA MASTER KONTIOLAHTI

Addresses

Location-aware search engine

A. Tabarcea, N. Gali and P. Fränti, "Framework for location-aware search engine", *Journal of Location Based Services*, 11 (1), 50-74, November 2017.



Clustering algorithms

Random Swap (RS)

Random Swap(X) $\rightarrow C, P$

$C \leftarrow$ Select random representatives(X);

$P \leftarrow$ Optimal partition(X, C);

REPEAT T times

$(C^{new}, j) \leftarrow$ Random swap(X, C);

$P^{new} \leftarrow$ Local repartition(X, C^{new}, P, j);

$C^{new}, P^{new} \leftarrow Kmeans(X, C^{new}, P^{new})$;

IF $f(C^{new}, P^{new}) < f(C, P)$ THEN

$(C, P) \leftarrow C^{new}, P^{new}$;

RETURN (C, P);

CI = 0

P. Fränti, "Efficiency of random swap clustering",
Journal of Big Data, 5:13, 1-29, 2018

Genetic Algorithm (GA)

GeneticAlgorithm(X) $\rightarrow (C, P)$

FOR $i \leftarrow 1$ TO Z DO

$C^i \leftarrow$ RandomCodebook(X);

$P^i \leftarrow$ OptimalPartition(X, C^i);

SortSolutions(C, P);

REPEAT

$\{C, P\} \leftarrow$ CreateNewSolutions($\{C, P\}$);

SortSolutions(C, P);

UNTIL no improvement;

CreateNewSolutions($\{C, P\}$) $\rightarrow \{C^{new}, P^{new}\}$

$C^{new-1}, P^{new-1} \leftarrow C^1, P^1$;

FOR $i \leftarrow 2$ TO Z DO

$(a, b) \leftarrow$ SelectNextPair;

$C^{new-i}, P^{new-i} \leftarrow$ Cross(C^a, P^a, C^b, P^b);

IterateK-Means(C^{new-i}, P^{new-i});

Cross(C^1, P^1, C^2, P^2) $\rightarrow (C^{new}, P^{new})$

$C^{new} \leftarrow$ CombineCentroids(C^1, C^2);

$P^{new} \leftarrow$ CombinePartitions(P^1, P^2);

$C^{new} \leftarrow$ UpdateCentroids(C^{new}, P^{new});

RemoveEmptyClusters(C^{new}, P^{new});

IS(C^{new}, P^{new});

CombineCentroids(C^1, C^2) $\rightarrow C^{new}$

$C^{new} \leftarrow C^1 \cup C^2$

CombinePartitions(C^{new}, P^1, P^2) $\rightarrow P^{new}$

FOR $i \leftarrow 1$ TO N DO

IF $\|x_i - c_{p_i^1}\|^2 \leq \|x_i - c_{p_i^2}\|^2$ THEN

$p_i^{new} \leftarrow p_i^1$

ELSE

$p_i^{new} \leftarrow p_i^2$

END-FOR

UpdateCentroids(C^1, C^2) $\rightarrow C^{new}$

FOR $j \leftarrow 1$ TO $|C^{new}|$ DO

$c_j^{new} \leftarrow$ CalculateCentroid(P^{new}, j);

CI = 0

P. Fränti, "Genetic algorithm with deterministic crossover for vector quantization", *Pattern Recognition Letters*, 2000.

Distance functions

Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^D \|x_i - y_i\|^2}$$

Dice coefficients:

$$d_{Dice}(x, y) = 1 - \frac{2 \cdot |x \cap y|}{|x| + |y|}$$

Edit distance:

Minimum number of edit operations
(insert, delete, substitute)
to transform string x to string y.

Strings divided into bi-grams:

String (5)
{st, tr, ri, **in**, **ng**}

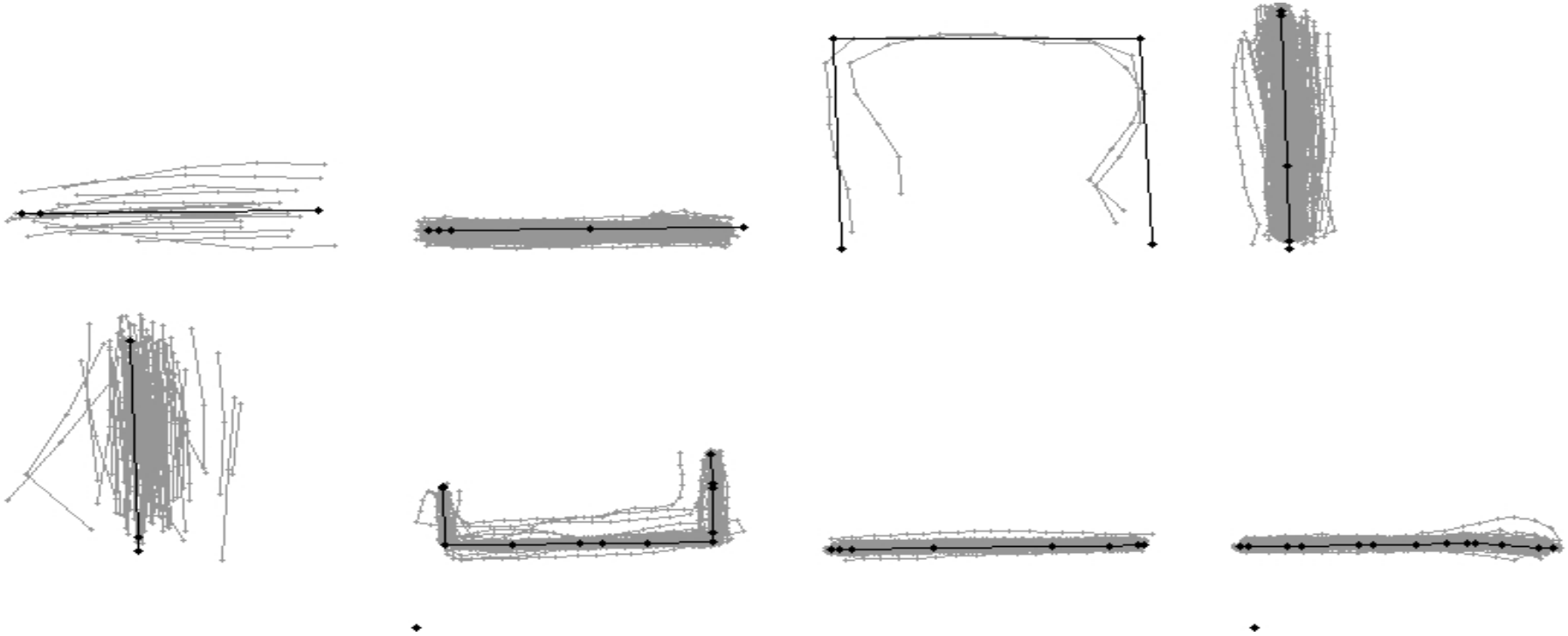
Mining (5)
{mi, **in**, ni, **in**, **ng**}

$x \cup y = \{st, tr, ri, in, ng, mi, ni\}$

$x \cap y = \{in, ng\}$

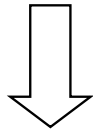
$d = 1 - 4 / (5 + 5) = 60\%$

Calculating mean



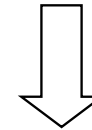
Means from scrambled text

HOELSVVKIG
HLSQPKPK QHSELINI
HELSSINI EVSDNFCKVM
THELSIFBJI EOMLSNI
HEHTLSINKI ZULSINKI
DHELSIRIWKJII



HELSINKI

REJOGNEITION
UCGNRTION RECOGUITPON
WEIOTNNIHTMIOJ
RRSCGNXIIUN
RRCOOGEPIONN RXONUIUOK
ECOBNFITUOND RUPCOWGNIPZTHUN
RCOGQNOIRTON



RECOGNITION

Clustering English words

466,544 words

Cluster 41	Cluster 43	Cluster 247	Cluster 292	Cluster 326
soft-bill soot-grimed sweet-toothed split-tongued black-visaged soft-winged short-witted short-termed stout-armed still-fishing stiff-limbed swift-stealing short-leaved snotty-nosed ivory-billed hot-mettled soft-going snowy-winged	Livingstone herringbone Burlingham Neowashingtonia Upington Hillingdon Lovington Arlington Lexington Herington Stringtown Arrington milliangstrom Accrington Northington Farlington Ellington	slommacky crummock bummack mimmock slammock bummalos mimmocky malmock hommocks earthgrubber crumhorn malbrouck krumhorn shammocky Babcock plumrock fleadock Cummock	Kurtz dinarchy myriarchy freshly triarcuated matriarch mandriarch dyarchic myriarch BSLArch taxiarch Bush Ruthi Knuth fleshy gush Thushi Furth	injelly johnin Conley moulvi Solly wolly Poulenc woody doozy oofy goodbyes coolly Woodlyn woofy boolya Lolly Coplay Goodbys

Clustering Tweets

544,113 tweets

If you're looking for work in #Oslo, Oslo, check out this #job: <https://t.co/ycYqTgJc8r> #BusinessMgmt #Hiring #CareerArc

If you're looking for work in #Göteborg, check out this #job: <https://t.co/bJGKlco2SN> #CustomerService #Hiring #CareerArc

Interested in a #job in #Uppsala, Uppsala County? This could be a great fit: <https://t.co/O7O91oVF1j> #Hiring #CareerArc

See our latest #kirkkonummi #job and click to apply: Software Developer Trainee - <https://t.co/vrsxjPMhsA> #SoftwareDev #Hiring #CareerArc

If you're looking for work in #Solna, check out this #job: <https://t.co/BUwsuBfXiO> #LEGO #Hiring #CareerArc

Många stör sig på andra på Twitter. På snälla Fredagen vill jag komma med lite bra tips: 1
<https://t.co/TFh40S7cC1...> <https://t.co/FjfR8ALvnn>

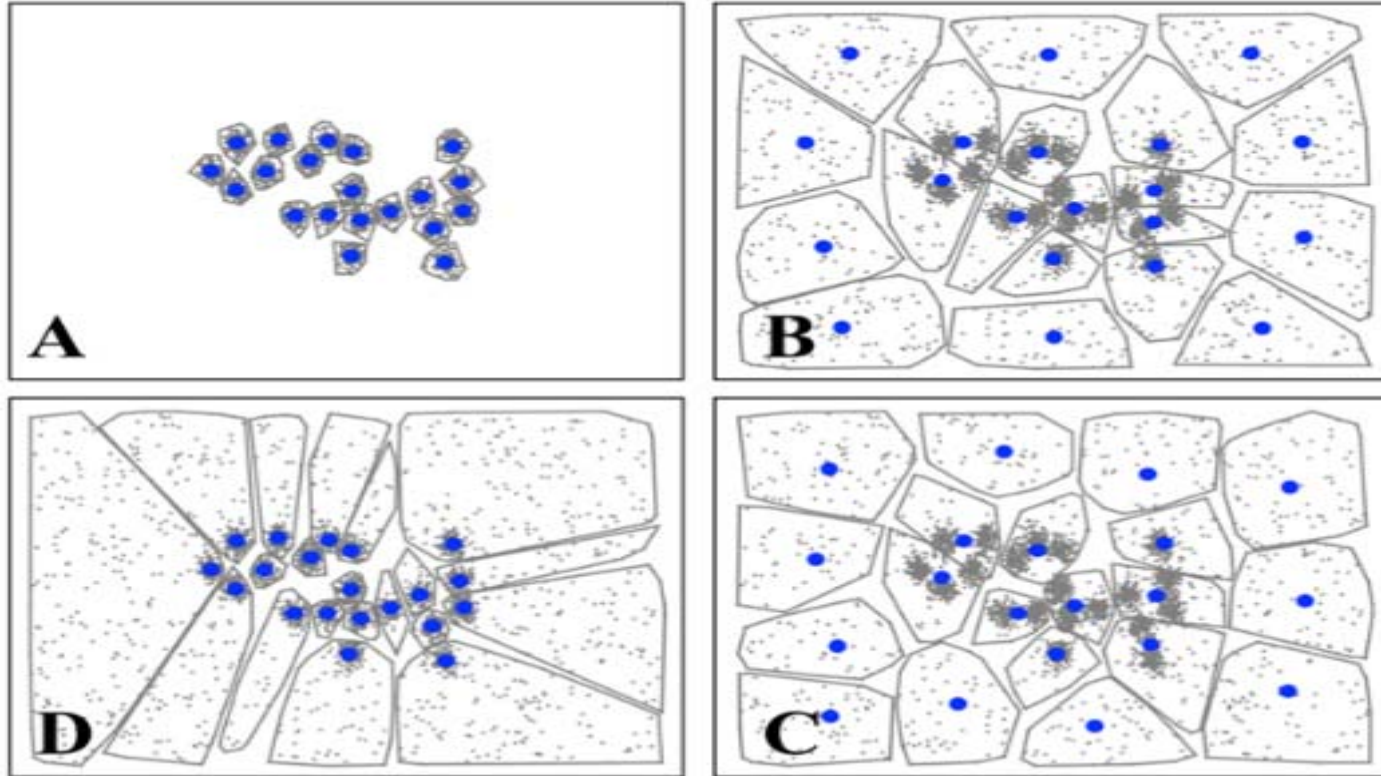
If you're looking for work in #HKI, check out this #job: <https://t.co/ZJLzylpqxf> #DellJobs #Sales #Hiring #CareerArc

See our latest #kirkkonummi #job and click to apply: SW Developer Intern, IoT Device and Data Management -...
<https://t.co/5GEkyiMUlh>

We're #hiring! Click to apply: Technical Program Manager - <https://t.co/BP0qrfigRK> #ProjectMgmt #stockholm #Job #Jobs #CareerArc



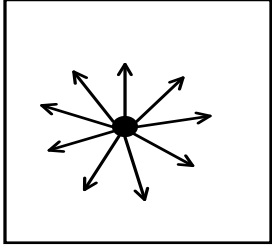
Detection and omission of outliers



Neighborhood graphs

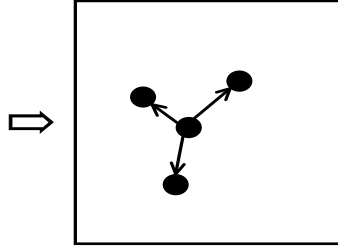
1. Fast NN-search

Full search:



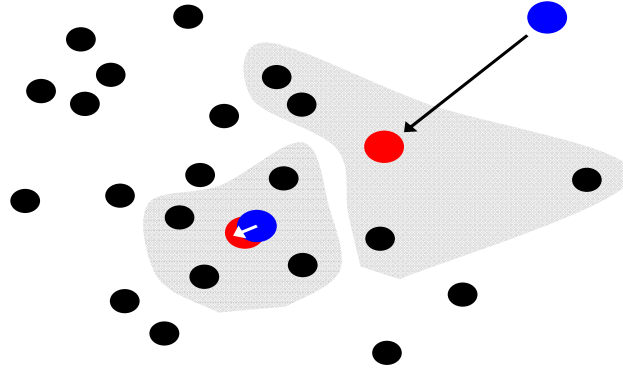
Full search:
 $O(M)$ distance
calculations.

Graph structure:

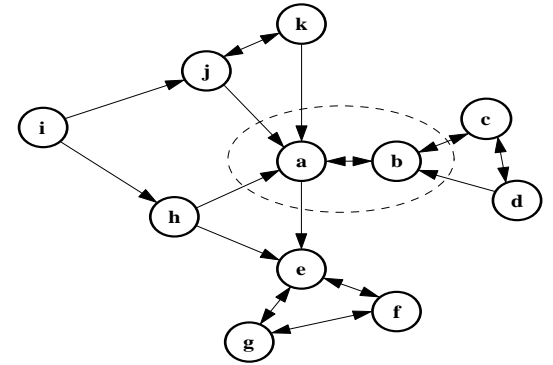


Graph structure:
 $O(k)$ distance
calculations.

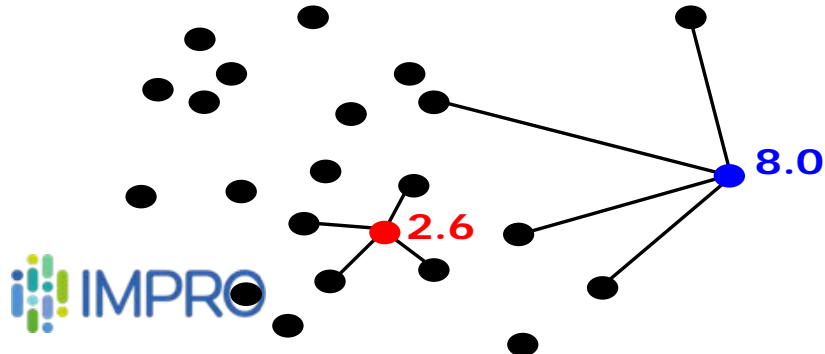
2. Outlier detection



3. Fast clustering

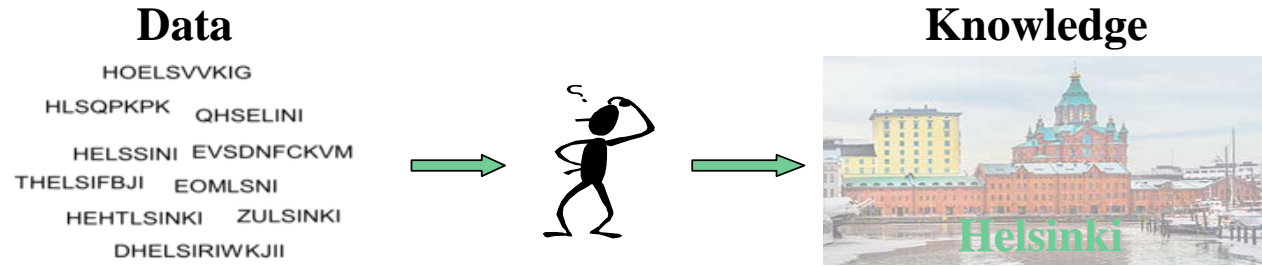


4. Density estimation



5. Recommendation system



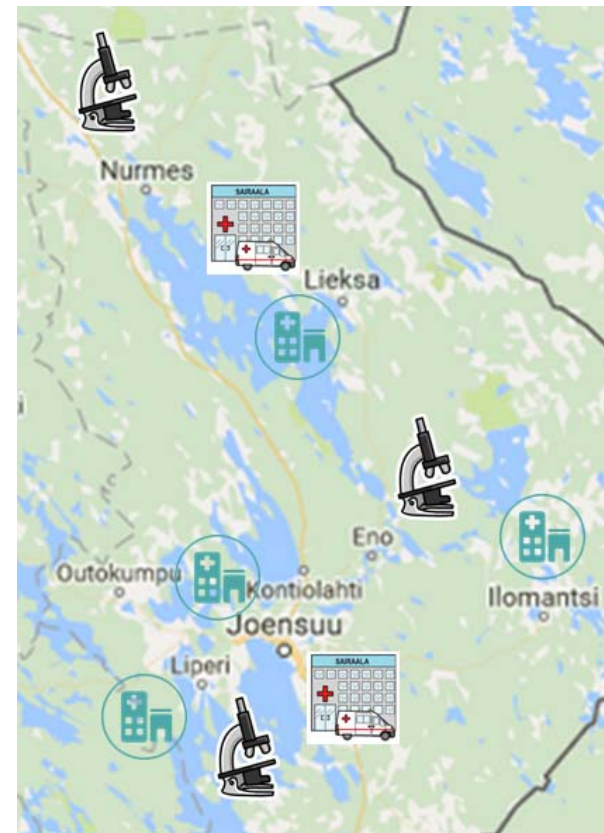


Health care services



Goals of the project

1. Create web interface
 - To access data
 - Importing external data in compatible formats
2. Interactive optimization tool
 - Working on map
 - Optimizes service locations and allocations for given cost provision site for specific groups
3. Algorithms and methods
 - Balanced data clustering
 - Network optimization
 - Data analysis



Where we started?



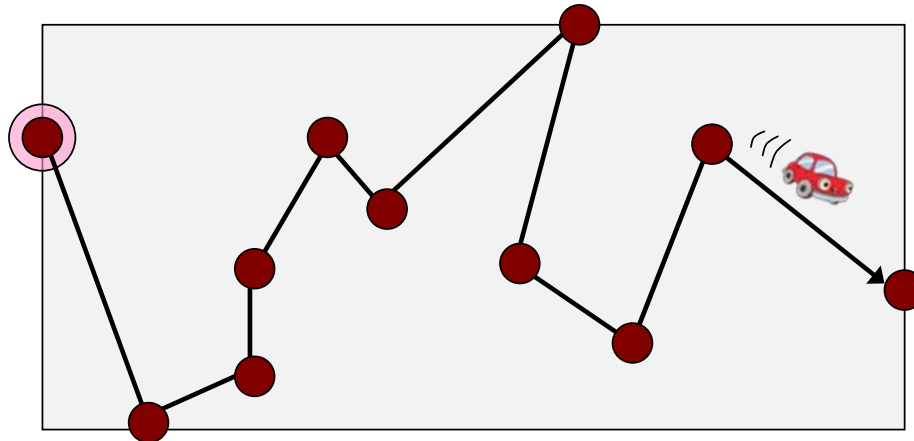
TSP

Clustering

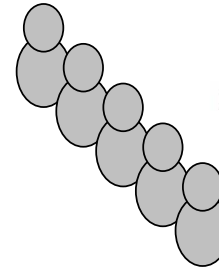
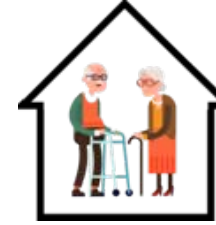
Work allocation

Scheduling

Touring nurse



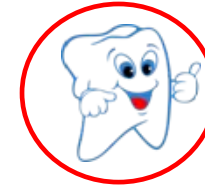
Dentist	8:00	FSDM	10:00
Physio	16:30		



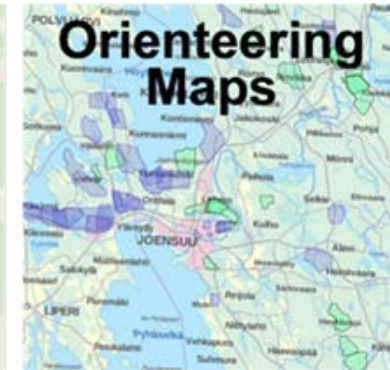


Health care services provided

1. Health center
2. Hospital
3. Dental
4. Rehabilitation
5. Aging therapies
6. Aging accomodation
7. Children's daycare

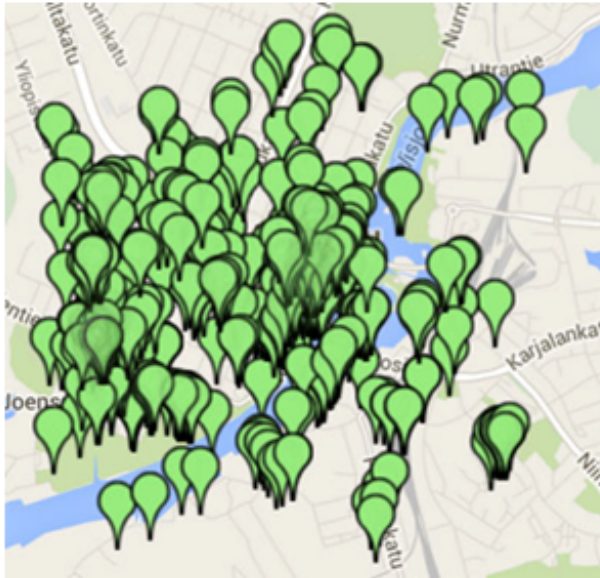


Solutions found on web

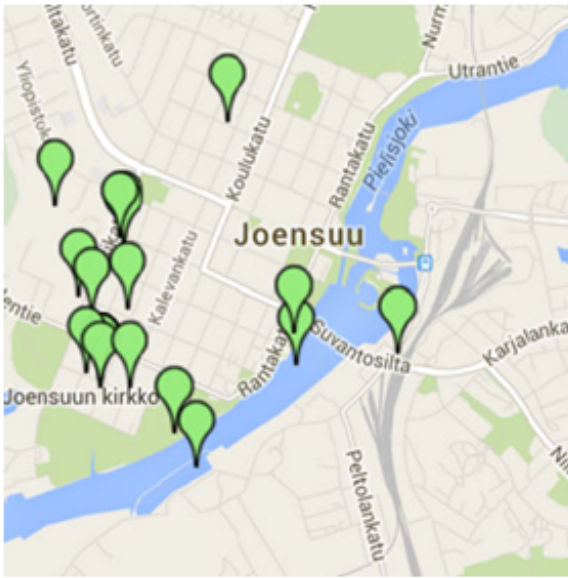


Clustering of data filtering?

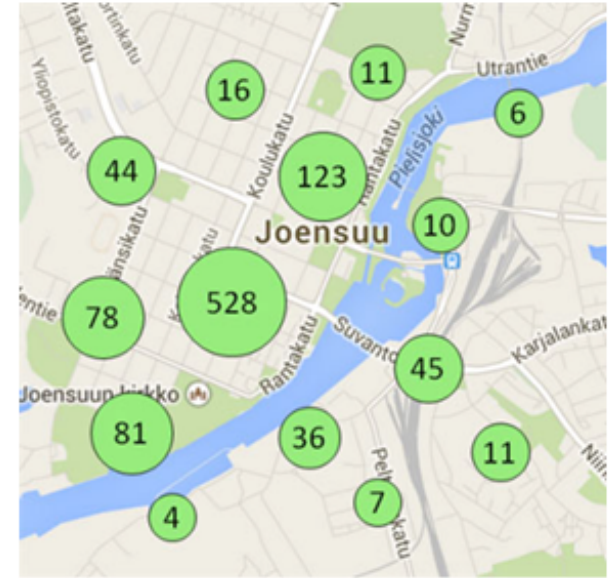
All data



Filtered data

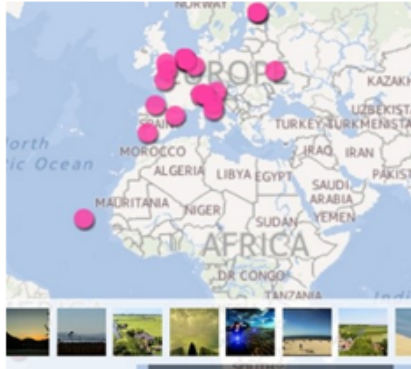


Clustered data



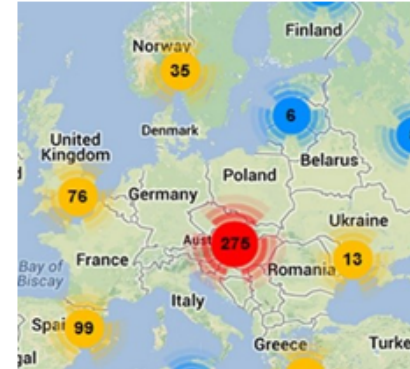
Visualizations of clustering (1)

Flickr



- (a) Filtering
- (b) Circle icon
- (c) -
- (d) -
- (e) -
- (f) Yes

Google maps



- (a) Clustering
- (b) Circle icon
- (c) Color, number
- (d) -
- (e) Yes
- (f) -

Panoramio



- (a) Filtering
- (b) Image icon
- (c) -
- (d) Yes
- (e) -
- (f) Yes

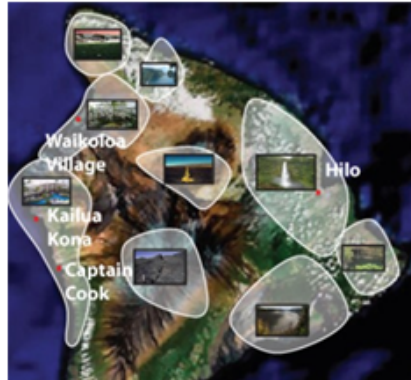
Voronoi [2]



- (a) Clustering
- (b) Voronoi
- (c) Color
- (d) Yes
- (e) Yes
- (f) -

Visualizations of clustering (2)

Container shape [11]



- (a) Clustering
- (b) Area, image
- (c) -
- (d) Yes
- (e) Yes
- (f) -

Mopsi



- (a) Clustering
- (b) Image icon
- (c) Number
- (d) -
- (e) Yes
- (f) Yes

Heat map [14]



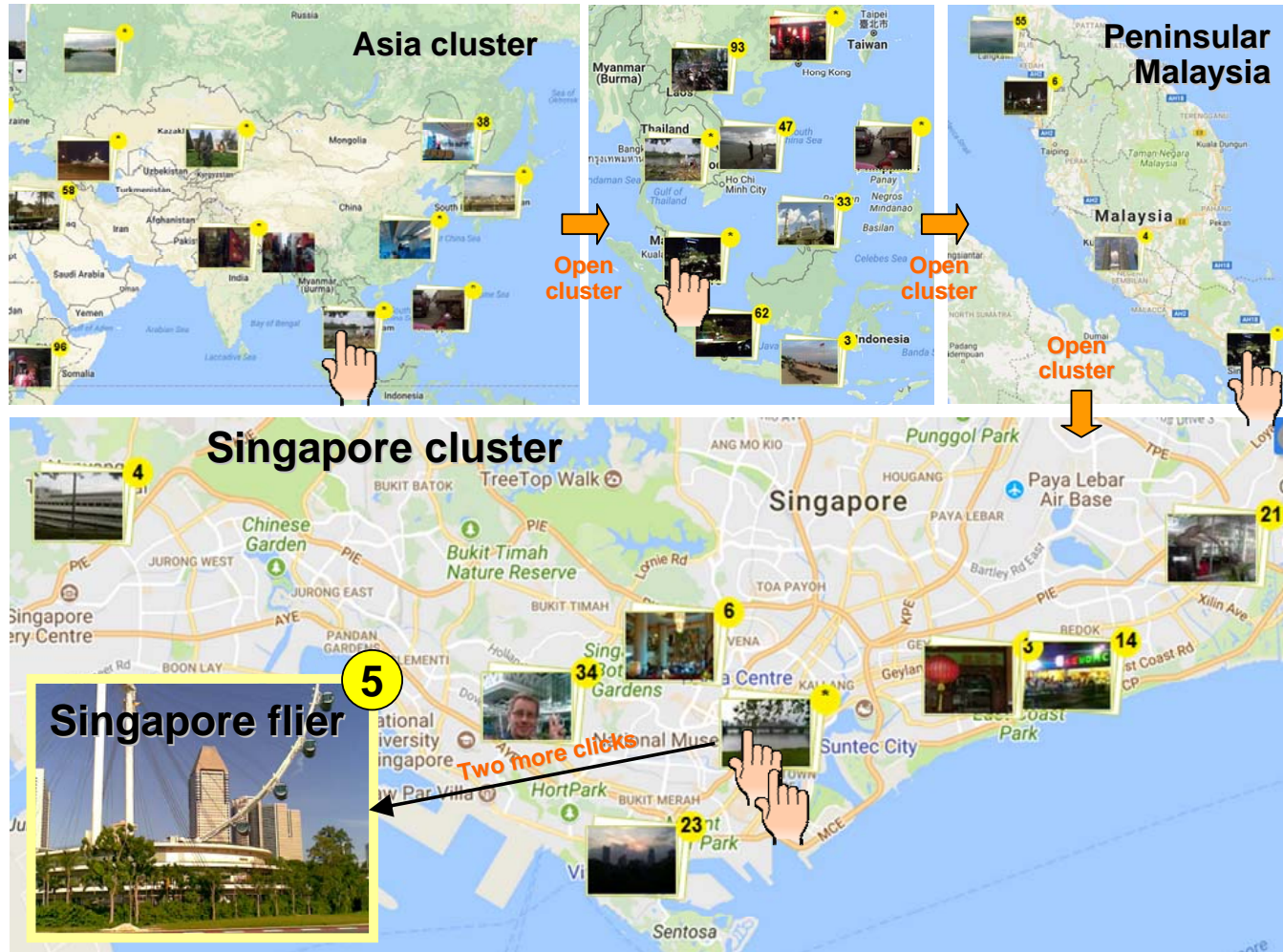
- (a) Clustering
- (b) Heat map
- (c) Color
- (d) Yes
- (e) Yes
- (f) -

Cells [12]



- (a) Clustering
- (b) Cell
- (c) Color
- (d) Yes
- (e) Yes
- (f) -

Clusters in Mopsi



Clusters in Mopsi

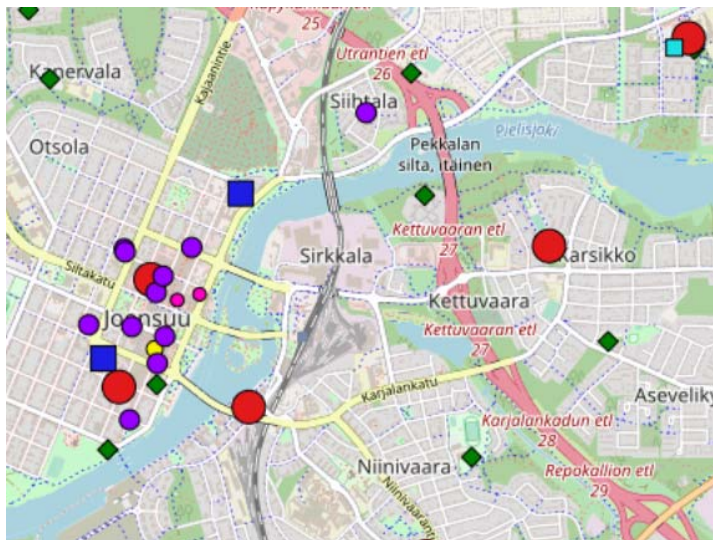


M. Rezaei and P. Fränti "Real-time clustering of large geo-referenced data for visualizing on map", *Advances in Electrical and Computer Engineering*, 2018.

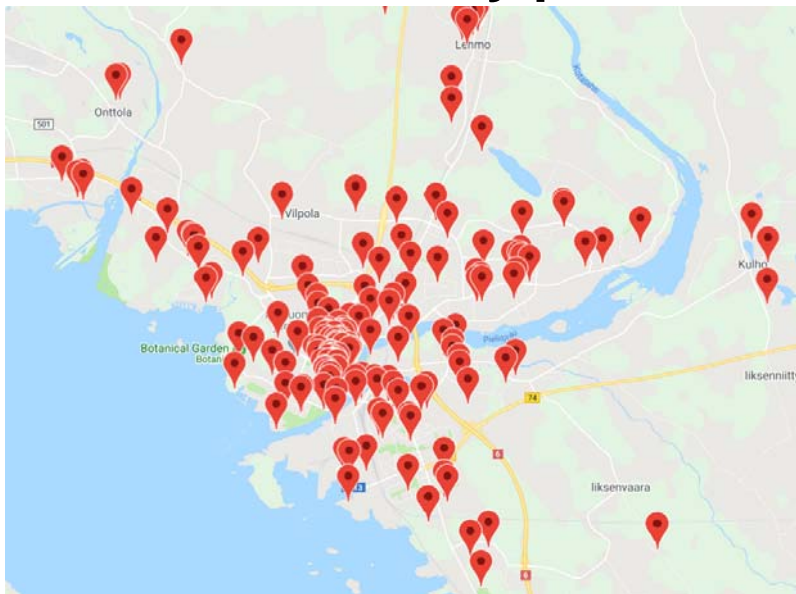


Various map view prototypes

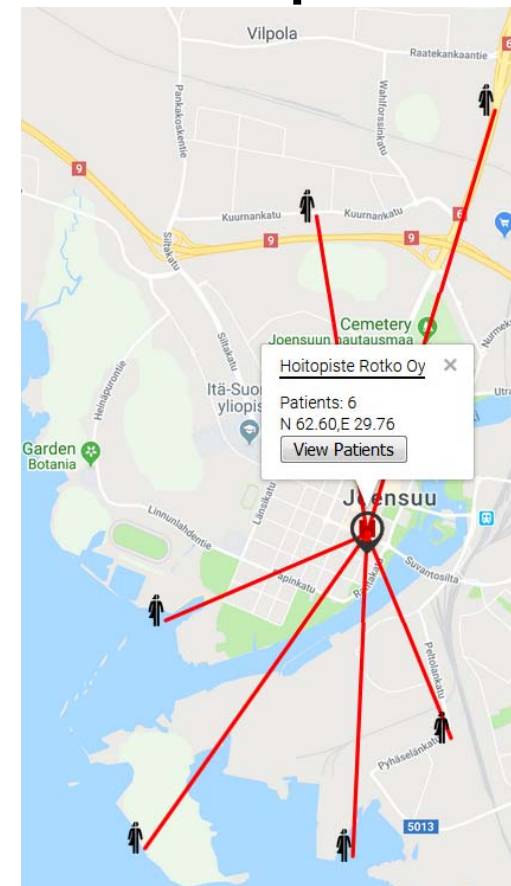
Geoserver image layers

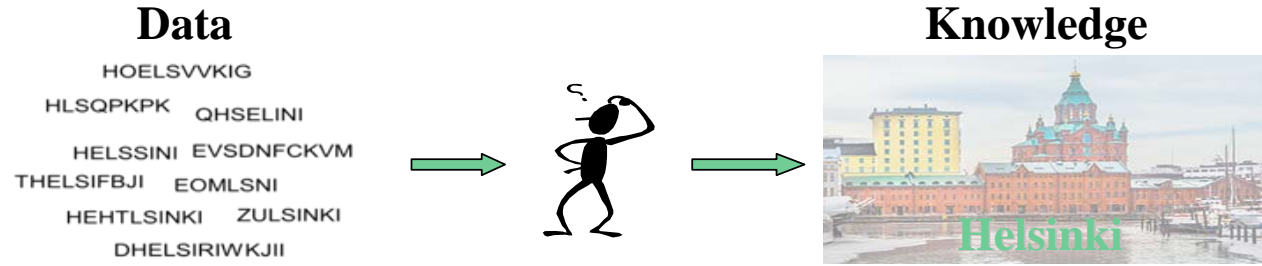


Quick-n-dirty pins



Team proto





Making the tools to meet the data

Correlation analysis of patient records

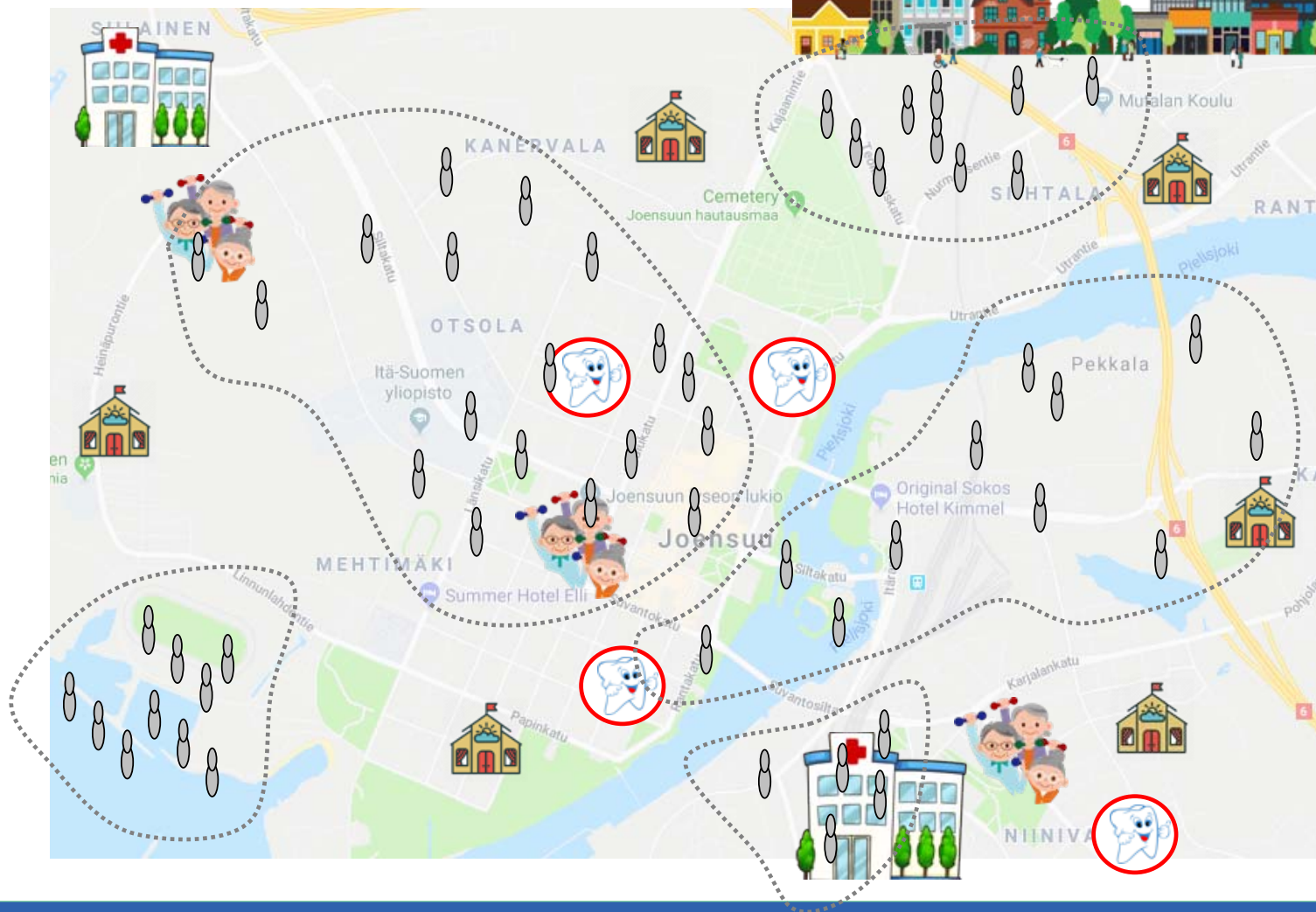
relation ▼	share(%) ◆	patients ◆	A ◆	B ◆	A	B
103.54	0.12	11	I85	K74	Esophageal varices	Fibrosis and cirrhosis of liver
22.98	0.14	13	C25	Z51	Malignant neoplasm of pancreas	Encounter for other aftercare and medical care
15.72	0.12	11	I35	I36	Nonrheumatic aortic valve disorders	Nonrheumatic tricuspid valve disorders
15.24	0.22	20	C34	Z51	Malignant neoplasm of bronchus and lung	Encounter for other aftercare and medical care
13.45	0.13	12	J44	J96	Other chronic obstructive pulmonary disease	Respiratory failure, not elsewhere classified
11.75	0.16	15	G47	J96	Sleep disorders	Respiratory failure, not elsewhere classified
10.84	0.12	11	H35	H43	Other retinal disorders	Disorders of vitreous body
10.79	0.16	15	K02	K04	Dental caries	Diseases of pulp and periapical tissues
10.66	0.14	13	H34	H40	Retinal vascular occlusions	Glaucoma
10.0	0.23	21	I34	I35	Nonrheumatic mitral valve disorders	Nonrheumatic aortic valve disorders

1. Esophageal varices – Fibrosis and Cirrhosis of liver
2. Malignant neoplasm of pancreas – Aftercare
3. Sleep disorders – Non-classified respiratory failure



Service locations in city

Clustering problem





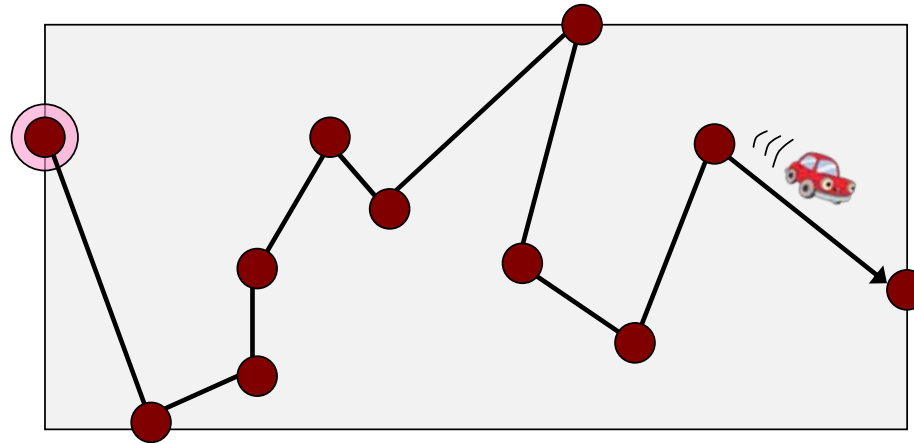
Services in rural area: touring nurse





Travelling salesman problem

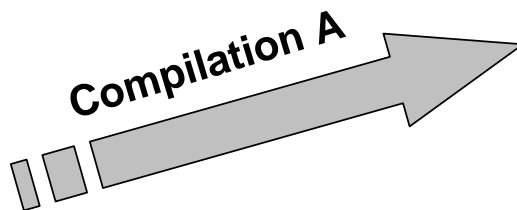
Touring nurse



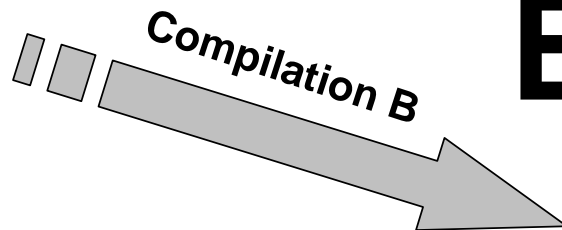
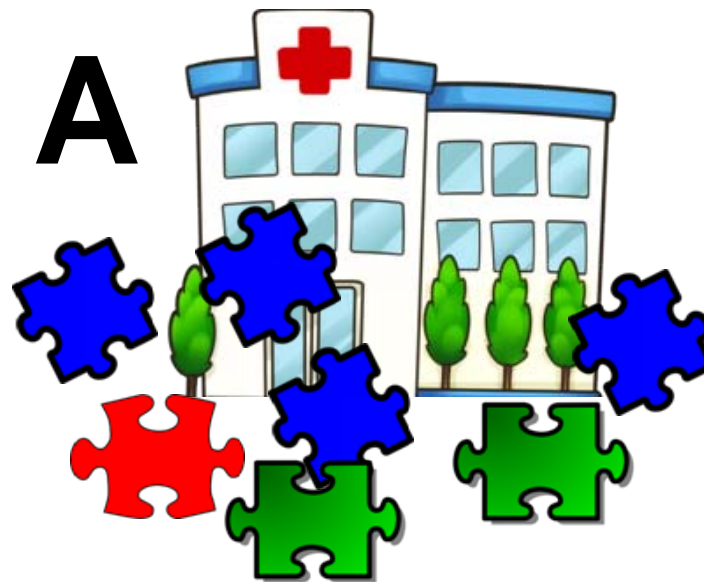


Recompiling the services

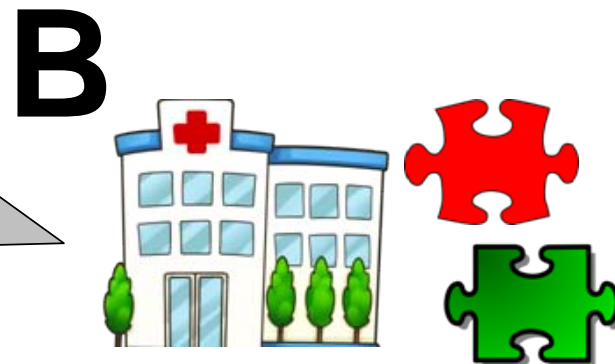
Resources



Compilation A

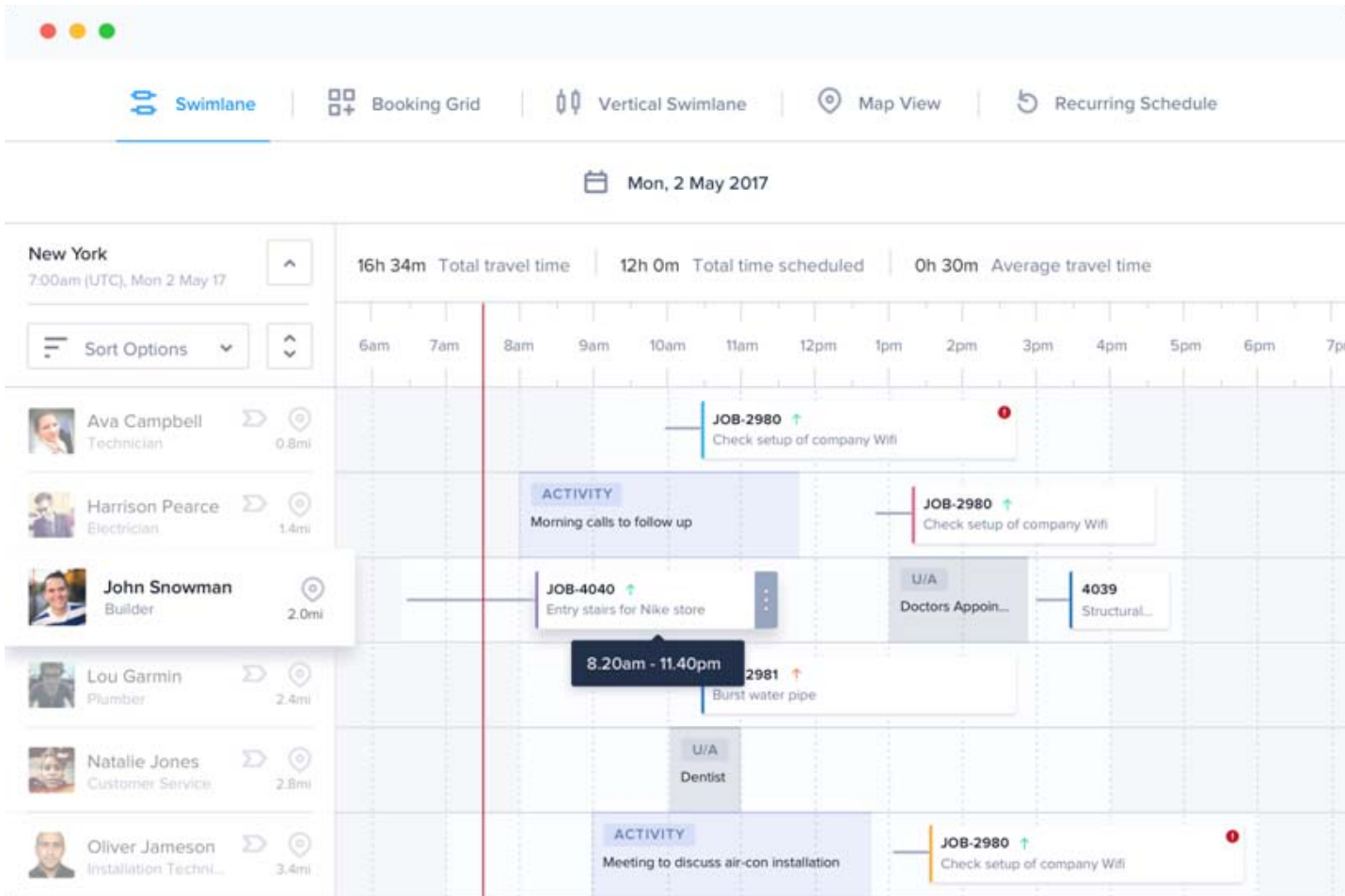


Compilation B





Scheduling problem



MOPSI project



Self-monitoring

Sleep management, maintain daily vitality

Monitor sleep, Bluetooth sync data to the phone and the cloud, so you know more about your quality of sleep



Portable heart rate expert

Combined with self-developed heart rate algorithm, real-time monitoring of heart rate during exercise to promote scientific exercise



Blood oxygen referred to as SPO2H, refers to the oxygen in the blood, the higher the content of blood oxygen, the better the human metabolism, normal blood oxygen saturation of 95% or more

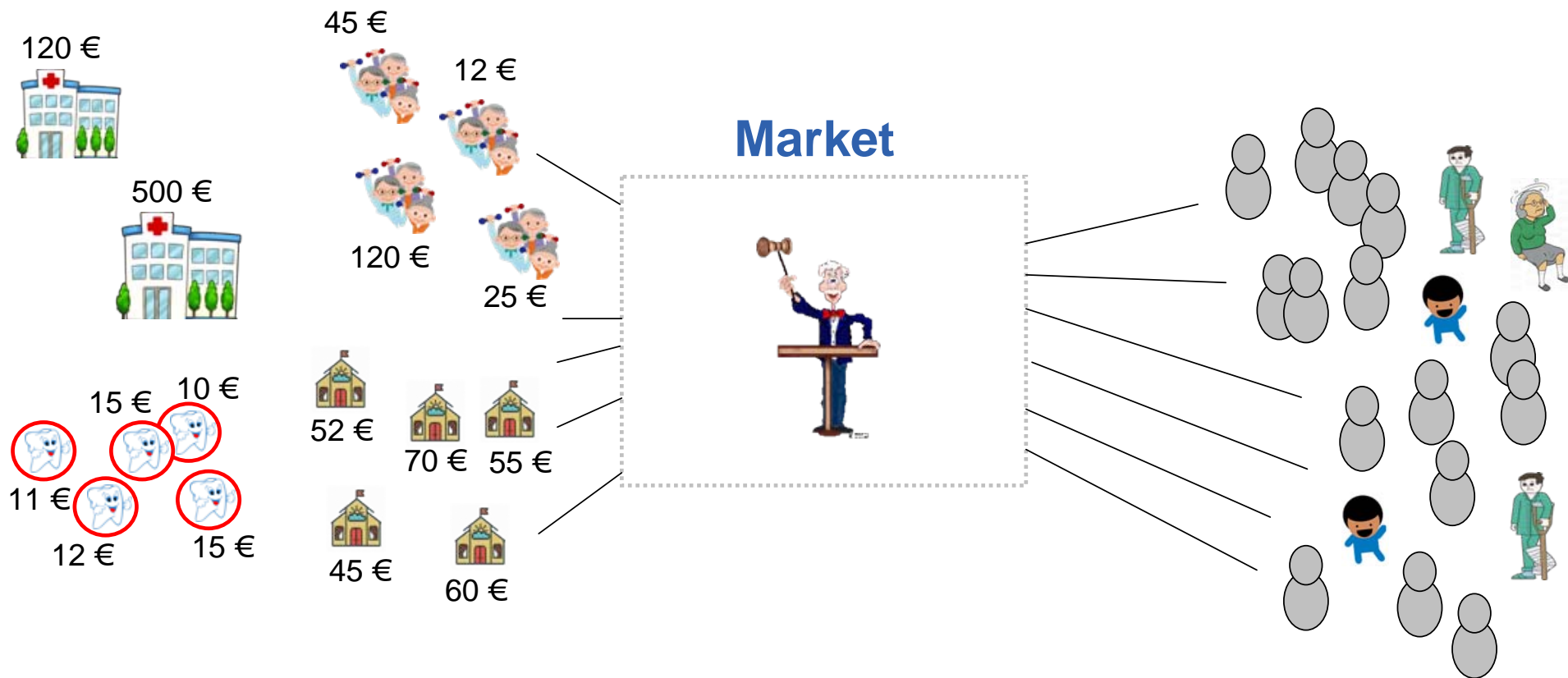




Market place of services

Ideal case

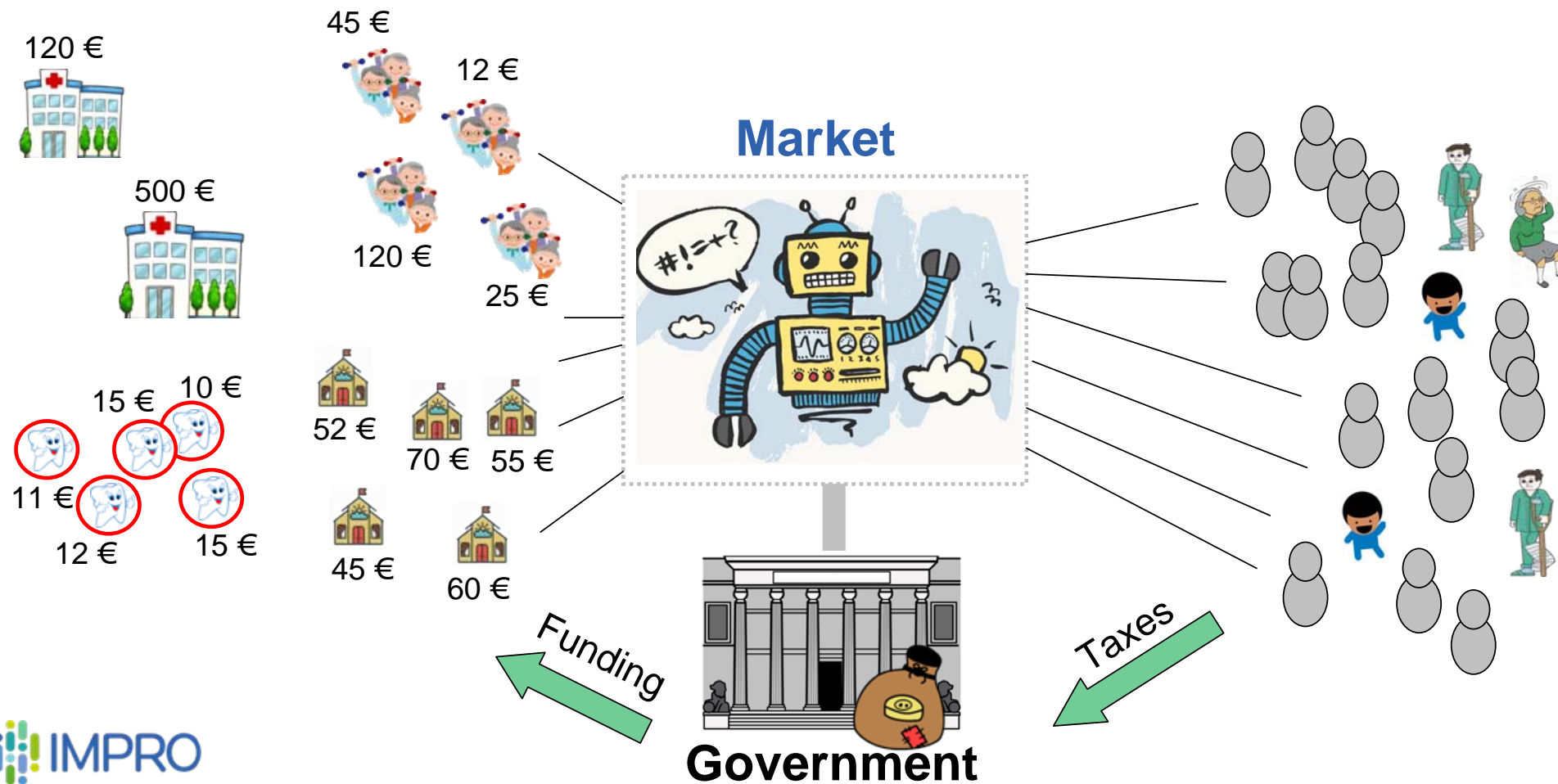
Market





Market place of services

Possible future scenario





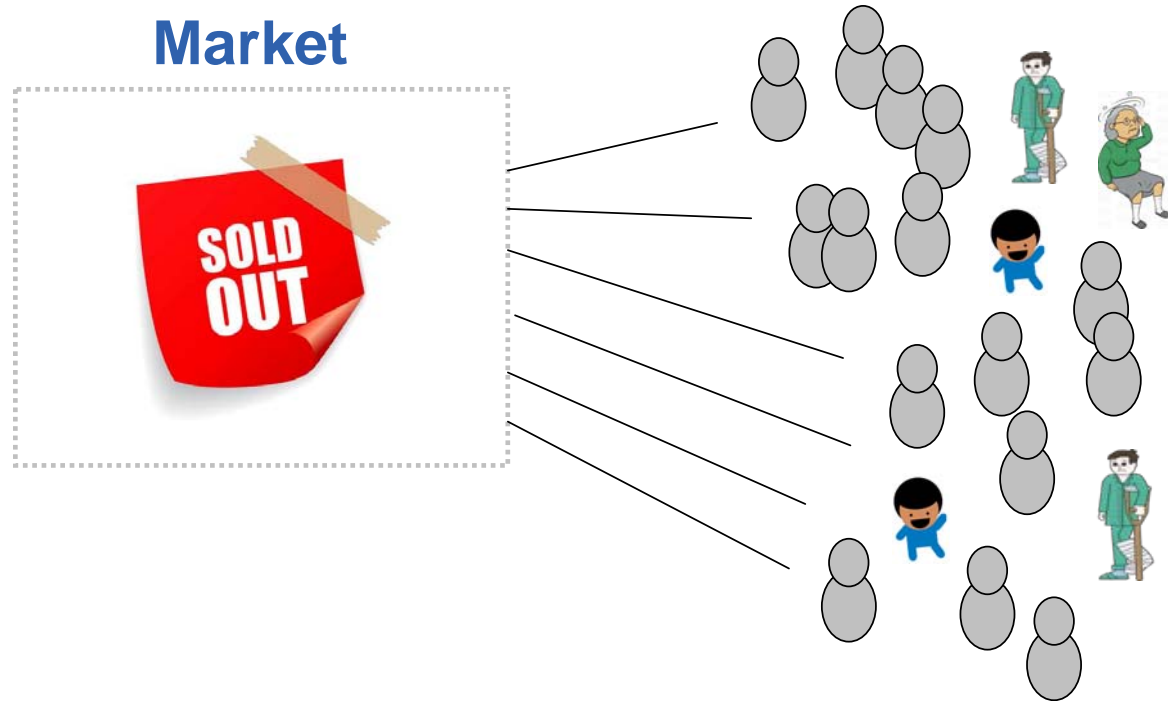
What to optimize exactly?

?



Case 1: No services, no costs

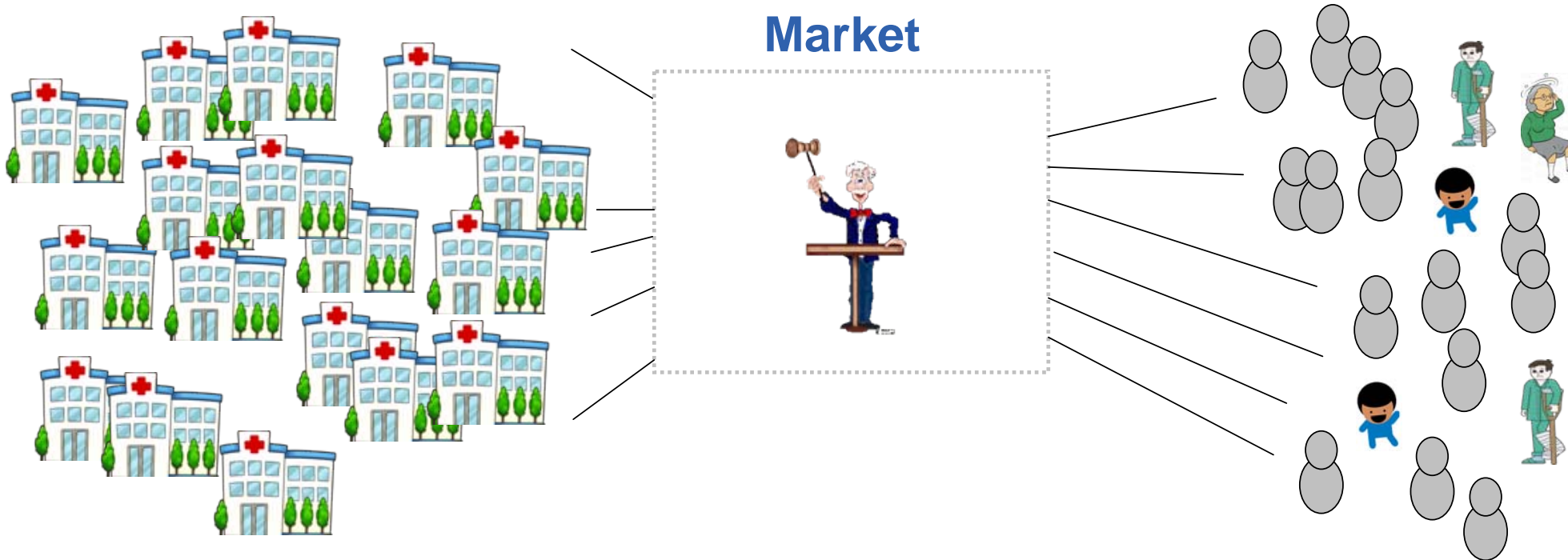
0 €





Case 2: Full service to everyone

1,000,000,000...€





The end

... future under process...