Christmas seminar 2018

UNIVERSITY OF EASTERN FINLAND

Social and health care services as an optimization problem

Pasi Fränti
Univ. of Eastern Finland
Joensuu, FINLAND

IMPRO

UNIVERSITY OF
EASTERN FINLAND

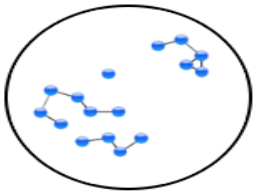
MACHINE LEARNING

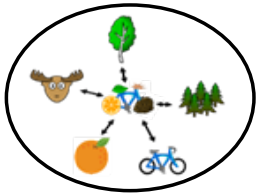
UNIVERSITY OF EASTERN FINLAND

Data mining
Information retrieval
Location-aware applications

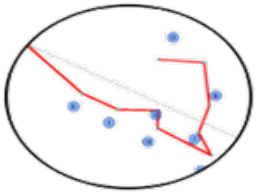
Location-aware applications

Clustering

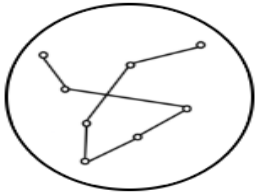




GPS trajectories

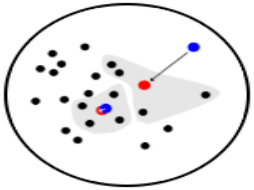


Spatial data

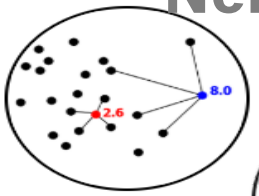


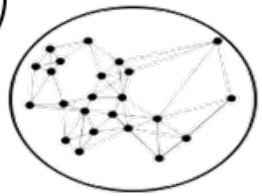
TSP

Games

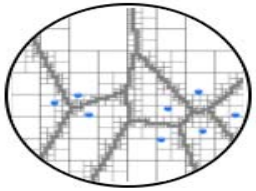


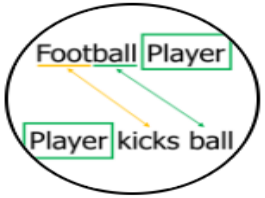
Neighborhood graphs





Clustering





Text mining

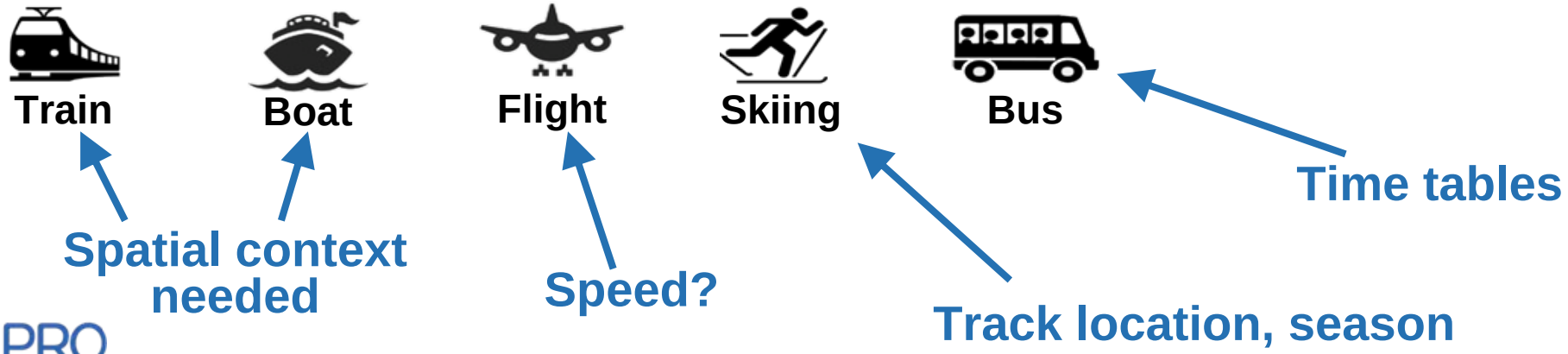
IMPRO

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:

Other possibilities:



IMPRO

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

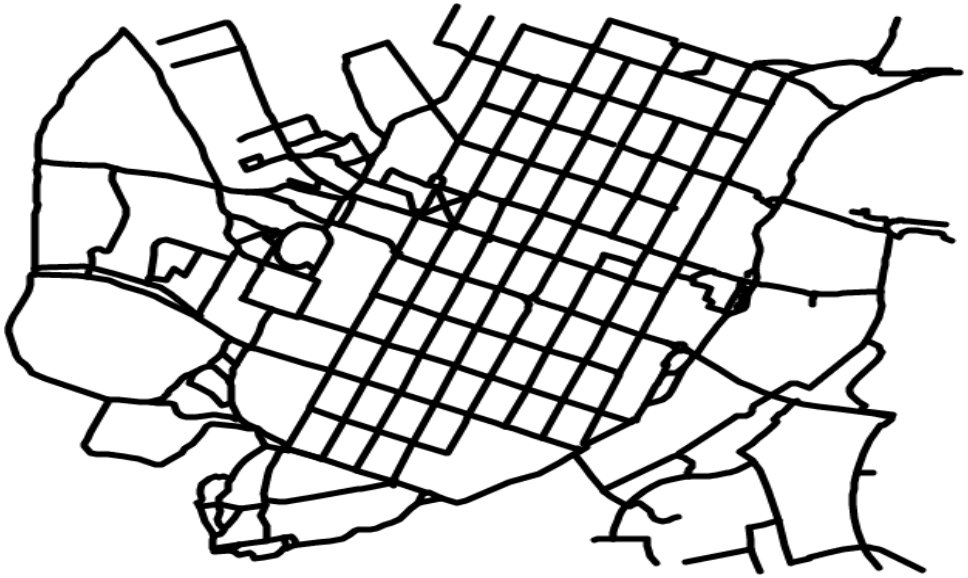
IMPRO

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

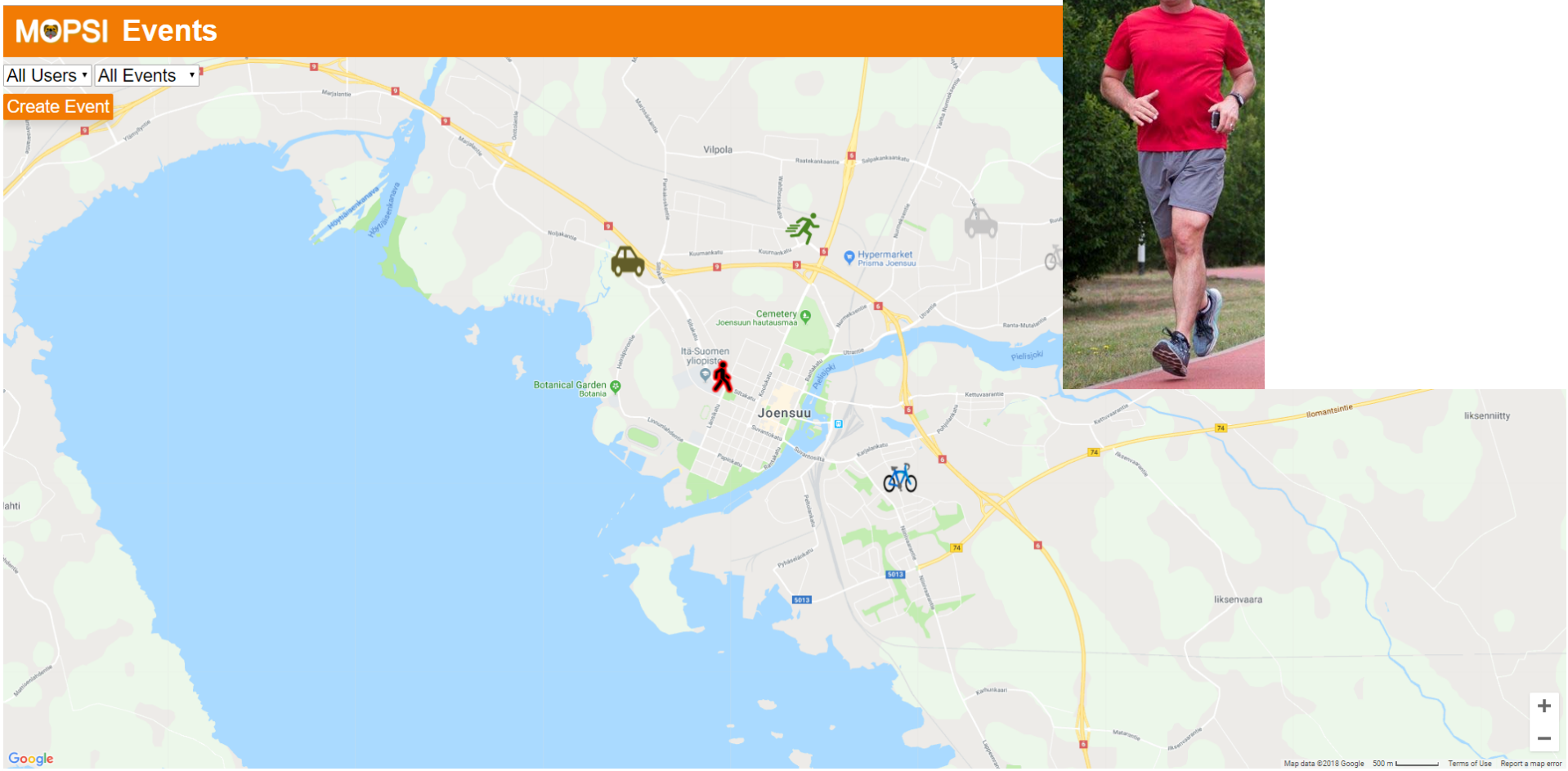
GPS Trajectories

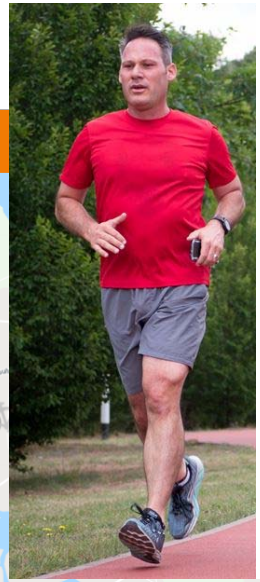
Road Network



IMPRO

Applications: share riding





Web content mining

Logo image

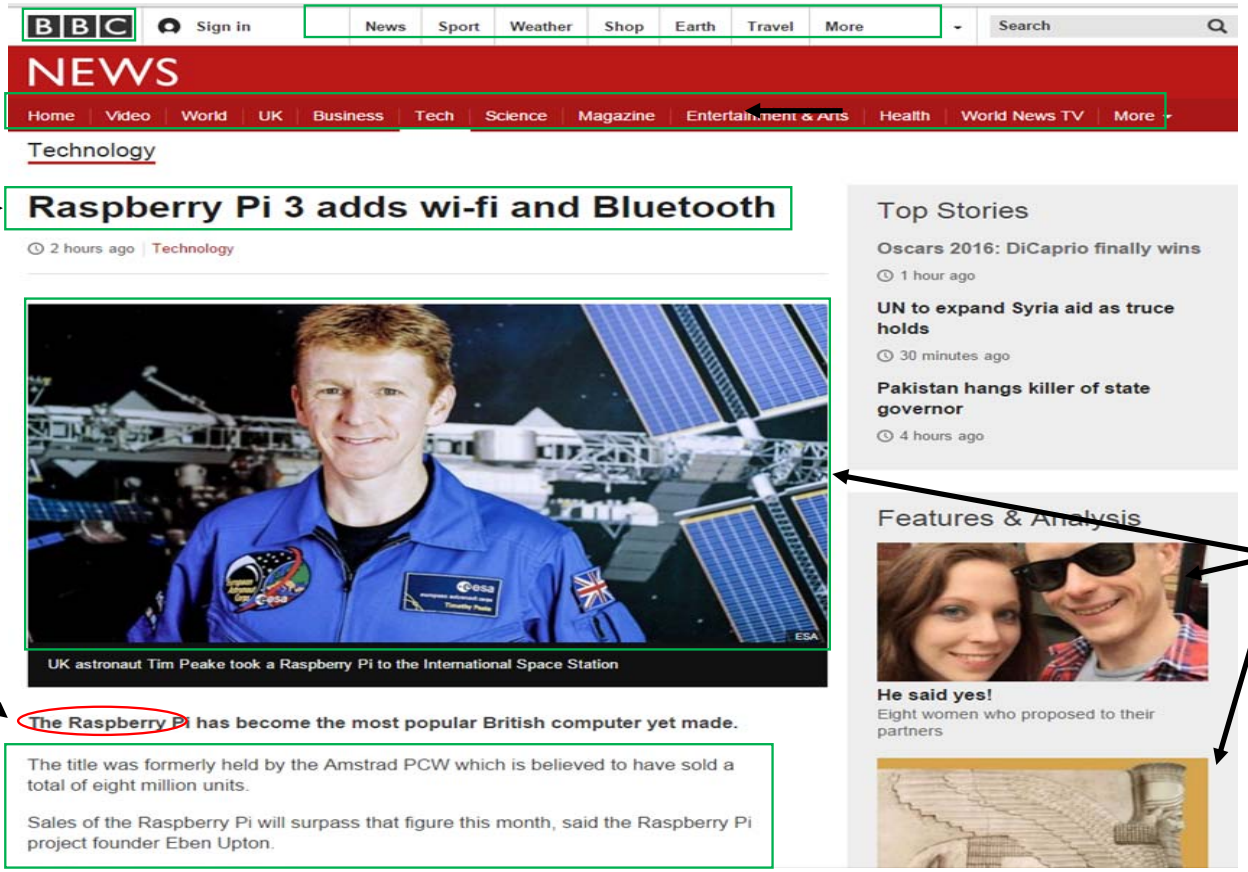
Title

Keywords

Text

Navigation bar

Images



Keyword extraction

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 Web Intelligence and Intelligent Agent Technology* Singapore, 2015

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

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.

IMPRO

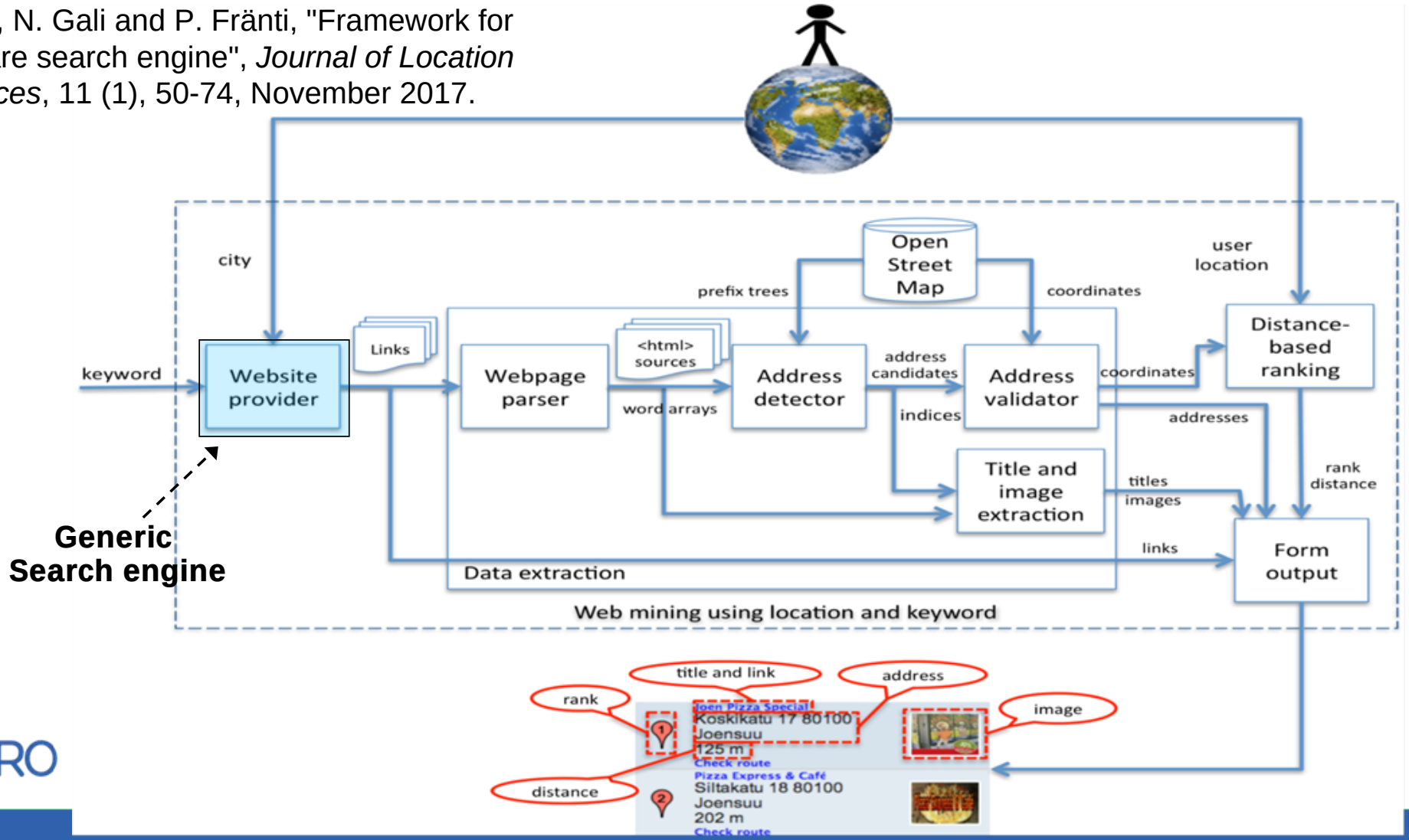
Address detection

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.

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} }
  $C^{new\text{-}1}$, $P^{new\text{-}1} \leftarrow C^1$, P^1;
  FOR $i \leftarrow$2 TO Z DO
    (a,b) $\leftarrow$ SelectNextPair;
    $C^{new\text{-}i}$, $P^{new\text{-}I} \leftarrow$ Cross(C^a, P^a, C^b, P^b);
    IterateK-Means($C^{new\text{-}i}$, $P^{new\text{-}i}$);

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 $\left\|x_i - c_{p_i^1}\right\|^2 \leq \left\|x_i - c_{p_i^2}\right\|^2$ 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.

IMPRO

Distance functions

Euclidean distance:

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

Edit distance:

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

Dice coefficients:

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

Strings divided into bi-grams:

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

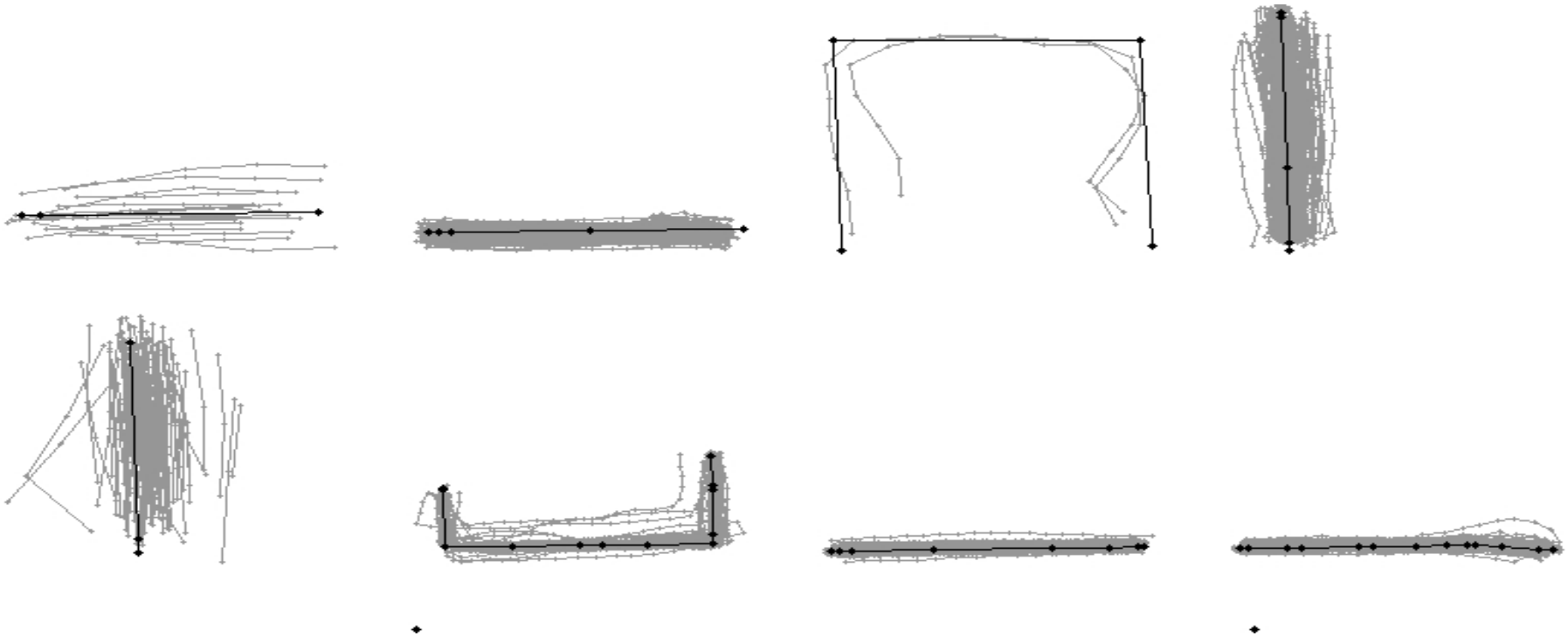
Mining (5)
{mi, **in**, ni, **in**, **ng**}

x∪y={st, tr, ri, in, ng, mi, ni}
x∩y={in, ng}

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

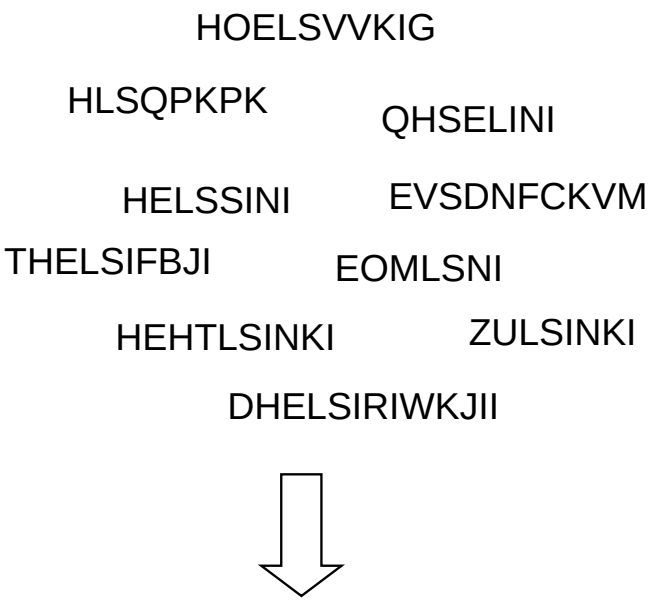
IMPRO

Calculating mean

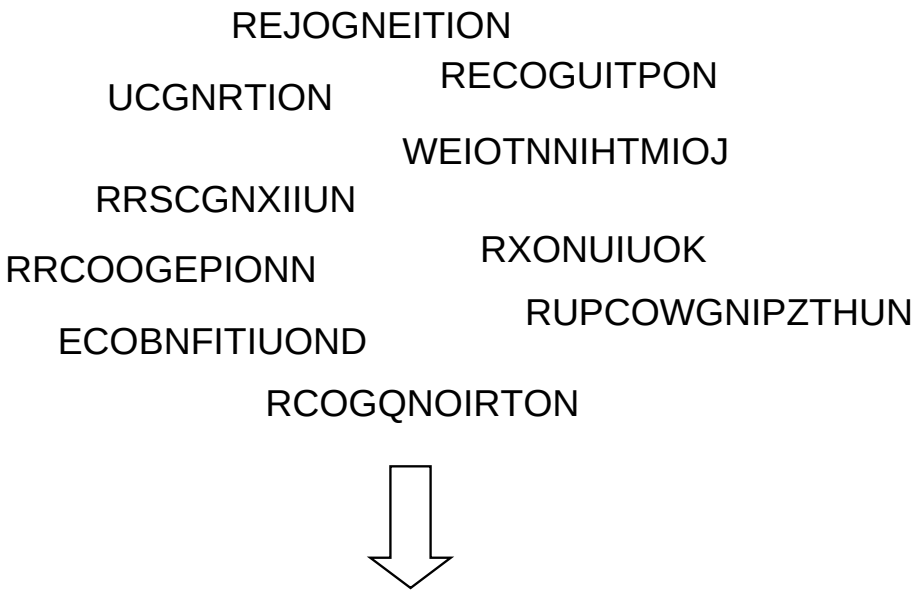


IMPRO

Means from scrambled text



HELSINKI



RECOGNITION

Clustering English words

466,544 words

Cluster 41	Cluster 43	Cluster 247	Cluster 292	Cluster 326
soft-bill	Livingstone	slommacky	Kurtz	injelly
soot-grimed	herringbone	crummock	dinarchy	johnin
sweet-toothed	Burlingham	bummack	myriarchy	Conley
split-tongued	Neowashingtonia	mimmock	freshly	moulvi
black-visaged	Upington	slammock	triarcuated	Solly
soft-winged	Hillingdon	bummalos	matriarch	wolly
short-witted	Lovington	mimmocky	mandriarch	Poulenc
short-termed	Arlington	malmock	dyarchic	woodsy
stout-armed	Lexington	hommocks	myriarch	doozy
still-fishing	Herington	earthgrubber	BSLArch	oofy
stiff-limbed	Stringtown	crumhorn	taxiarch	goodbyes
swift-stealing	Arrington	malbrouck	Bush	coolly
short-leaved	milliangstrom	krumhorn	Ruthi	Woodlyn
snotty-nosed	Accrington	shammocky	Knuth	woofy
ivory-billed	Northington	Babcock	fleshy	boolya
hot-mettled	Farlington	plumrock	gush	Lolly
soft-going	Ellington	fleadock	Thushi	Coplay
snowy-winged		Cummock	Furth	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/bJGKIco2SN #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

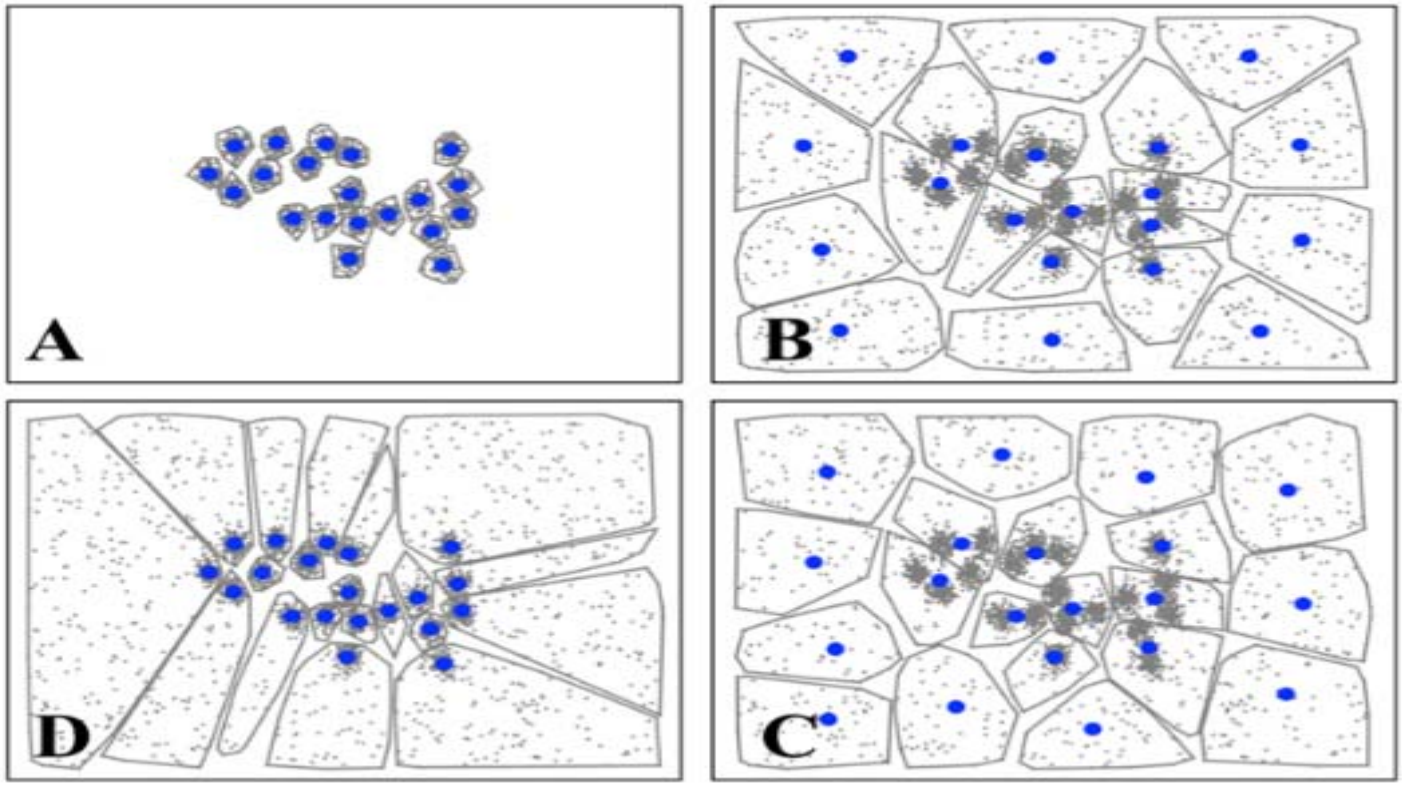
If you're looking for work in #HKI, check out this #job: https://t.co/ZJLzylpqxf #DellJobs #Sales #Hiring #CareerArc

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

IMPRO

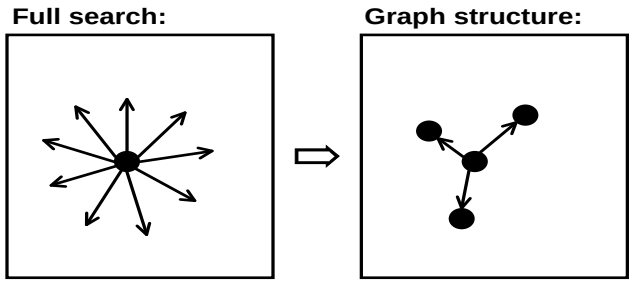
Detection and omission of outliers



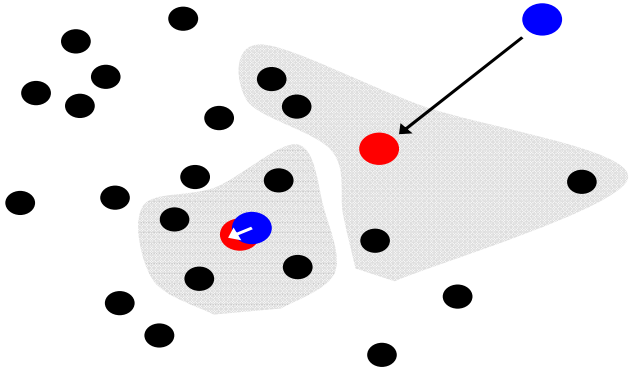
IMPRO

Neighborhood graphs

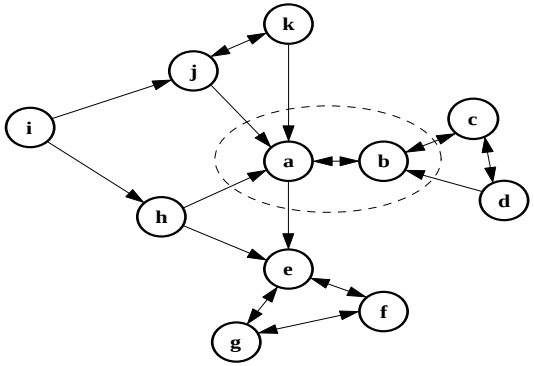
1. Fast NN-search



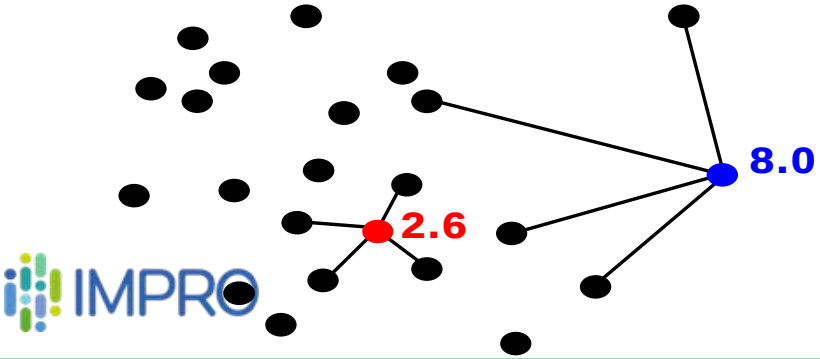
2. Outlier detection



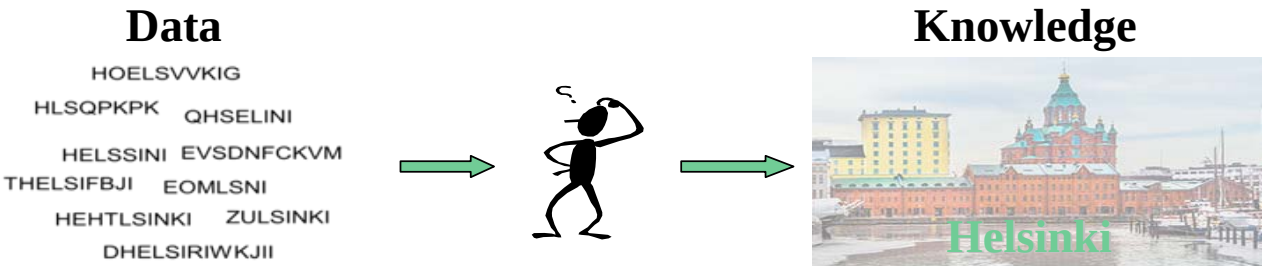
3. Fast clustering



4. Density estimation



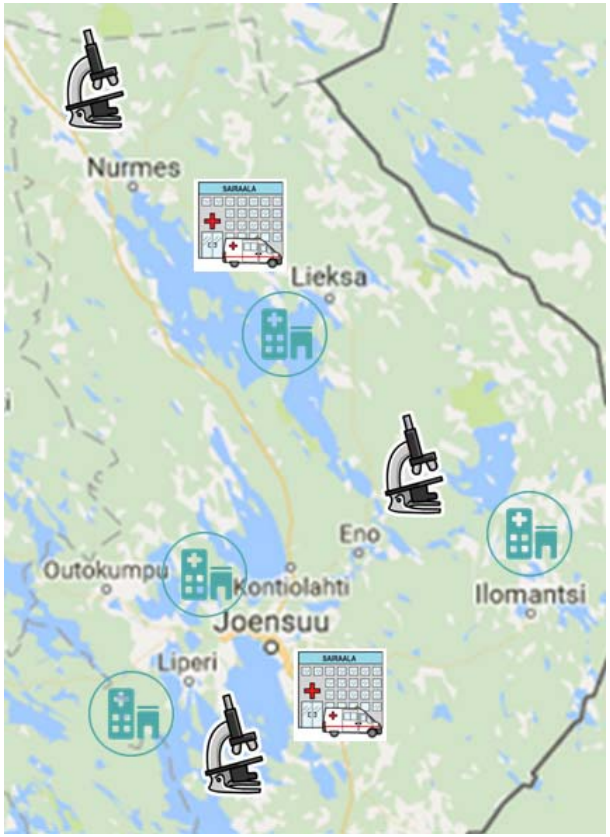
5. Recommendation system



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



IMPRO

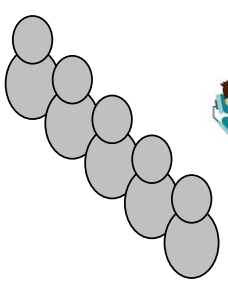
Where we started?

TSP

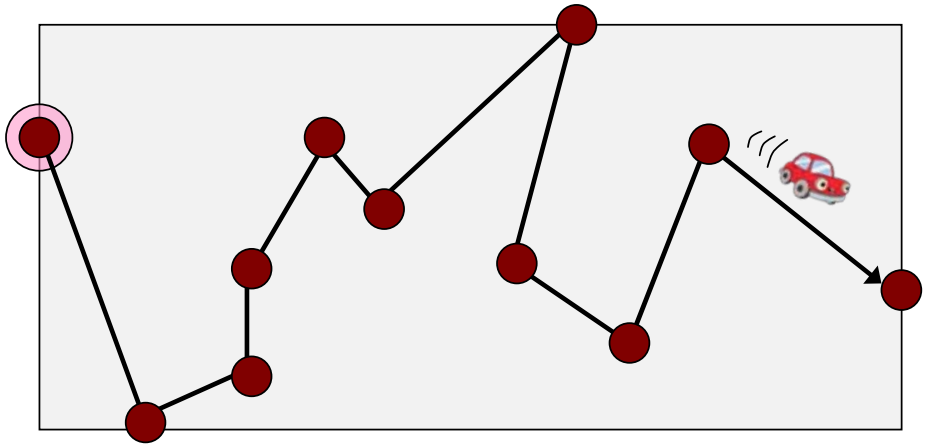
Clustering

Work allocation

Scheduling



Touring nurse



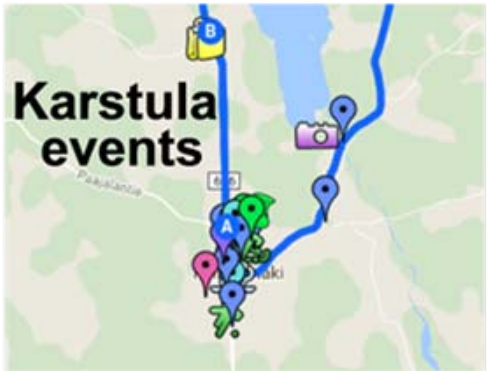
IMPRO

Health care services provided

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

IMPRO

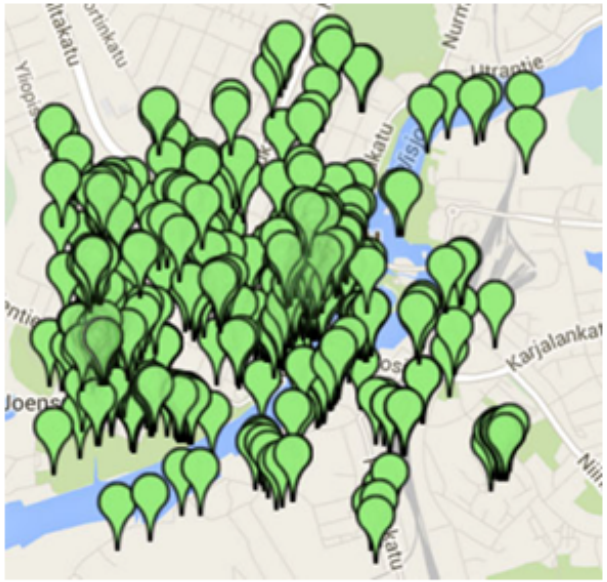
Solutions found on web



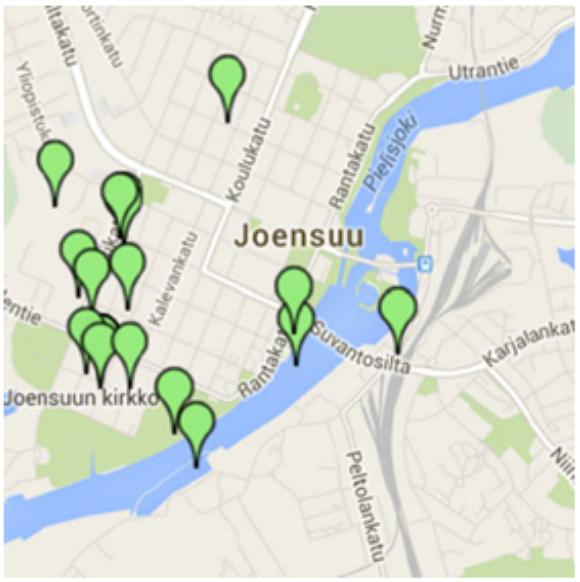
IMPRO

Clustering of data filtering?

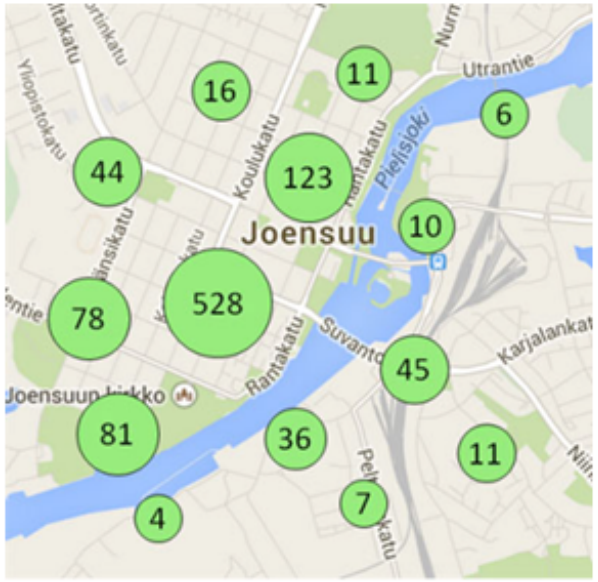
All data



Filtered data



Clustered data



IMPRO

Visualizations of clustering (1)

Flickr

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

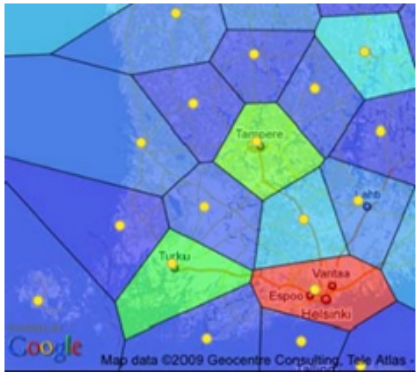
Google maps

(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]

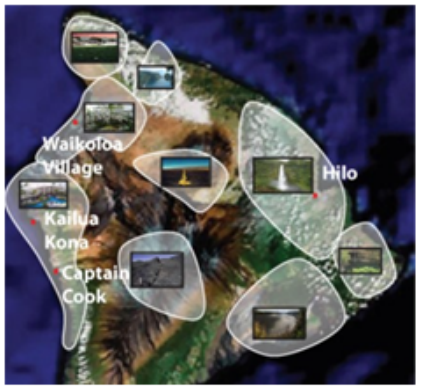


(a) Clustering
(b) Voronoi
(c) Color
(d) Yes
(e) Yes
(f) -

IMPRO

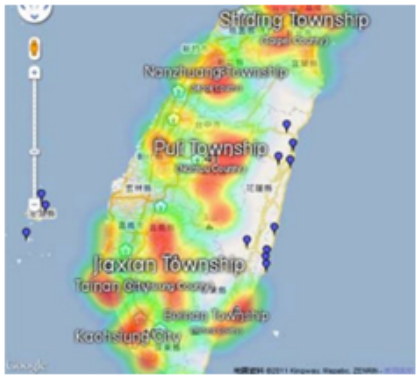
Visualizations of clustering (2)

Container shape [11]



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

Heat map [14]



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

Mopsi

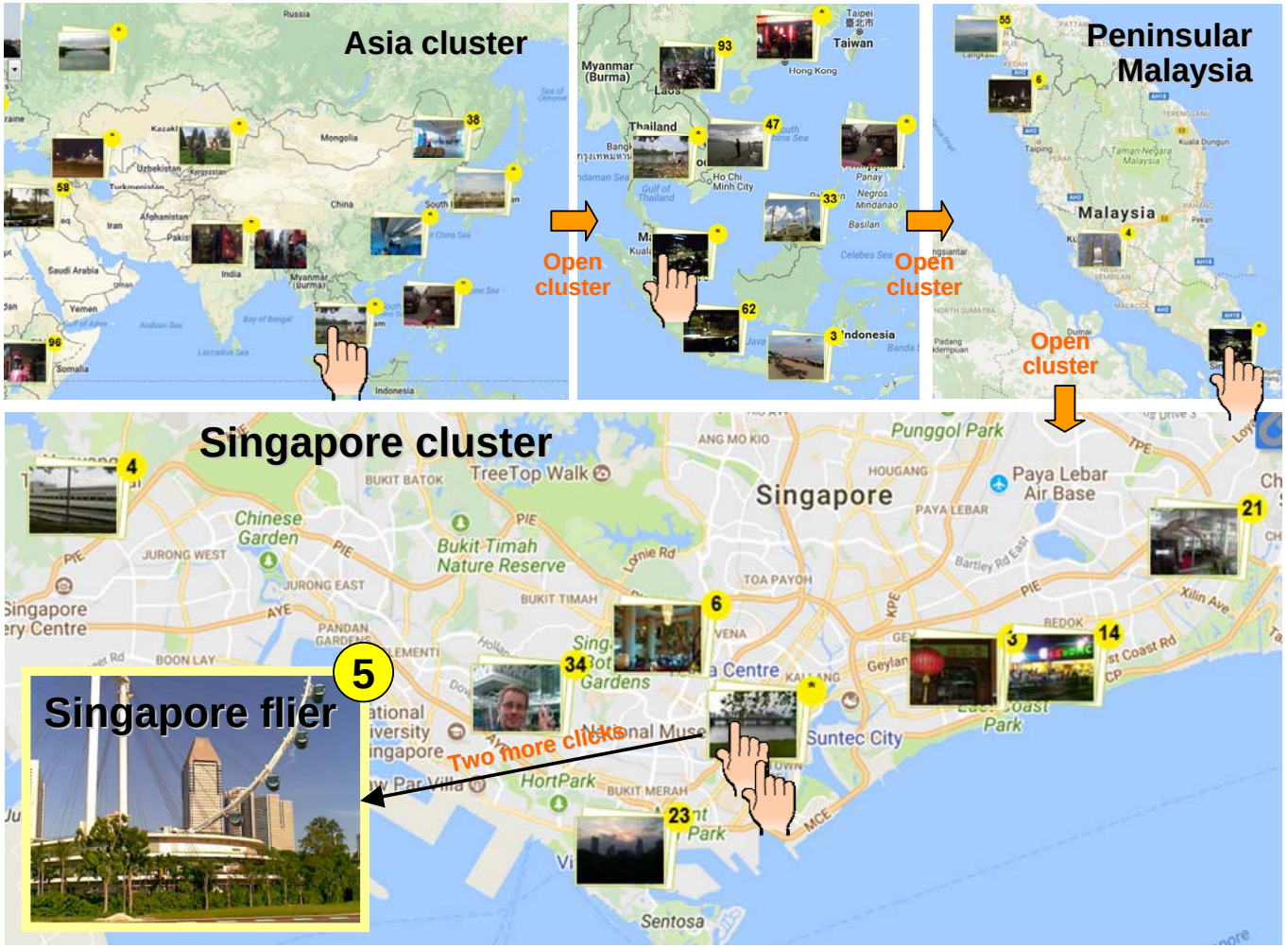
(a) Clustering
(b) Image icon
(c) Number
(d) -
(e) Yes
(f) Yes

Cells [12]

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

IMPRO

Clusters in Mopsi

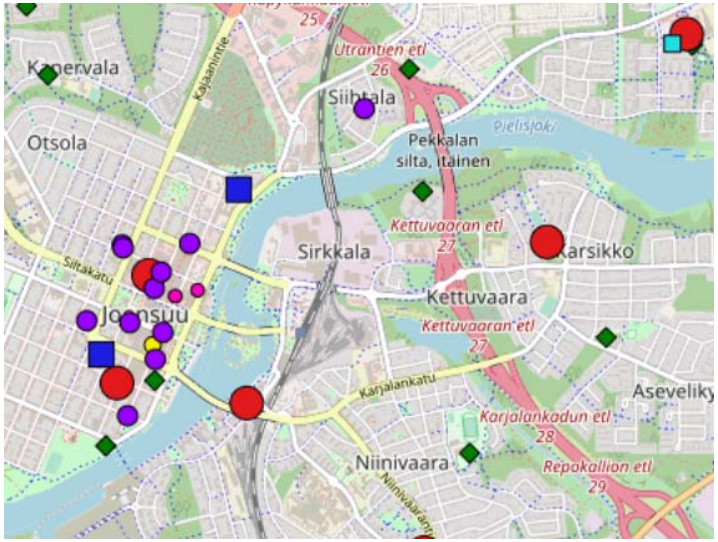


Clusters in Mopsi

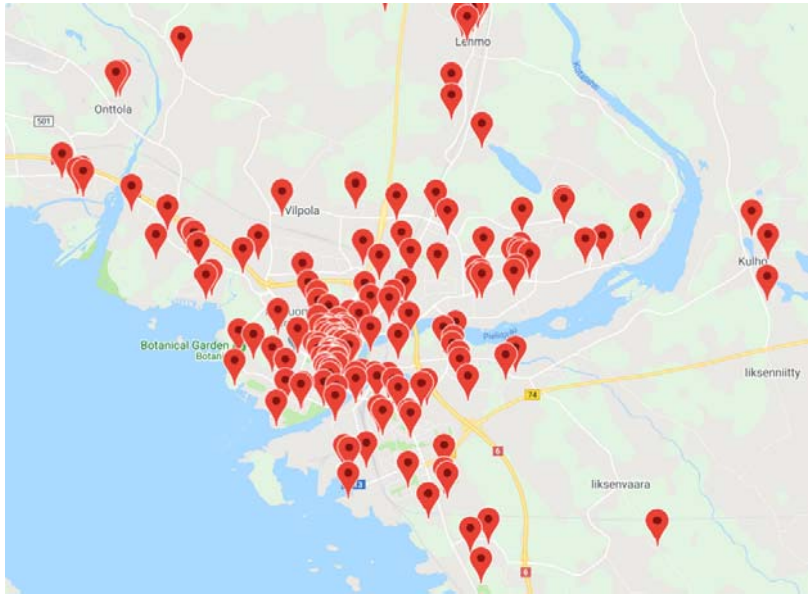
M. Rezaei and P. Fränti "Real-time clustering of large geo-referenced data for visualizing on map", *Advances in Electrical and Computer Engineering*, 2018.

Various map view prototypes

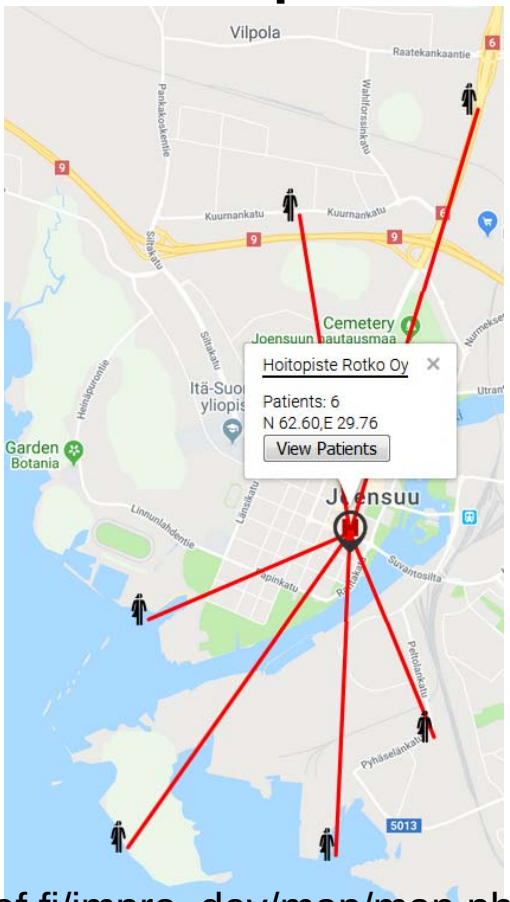
Geoserver image layers



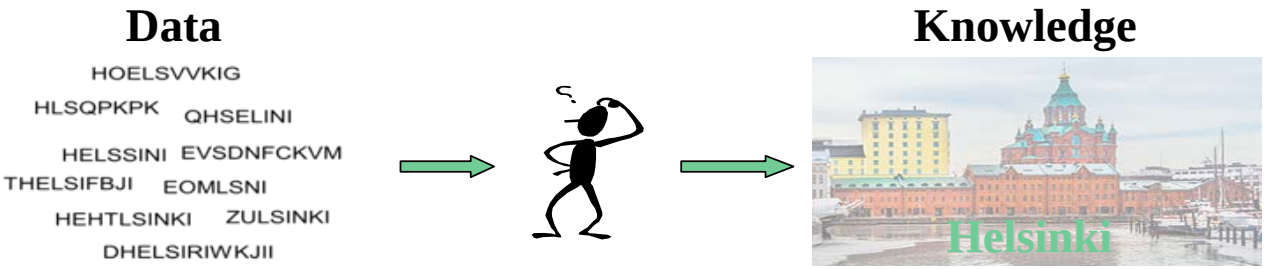
Quick-n-dirty pins



Team proto



http://cs.uef.fi/impro_dev/map/map.php



Making the tools to meet the data

IMPRO

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

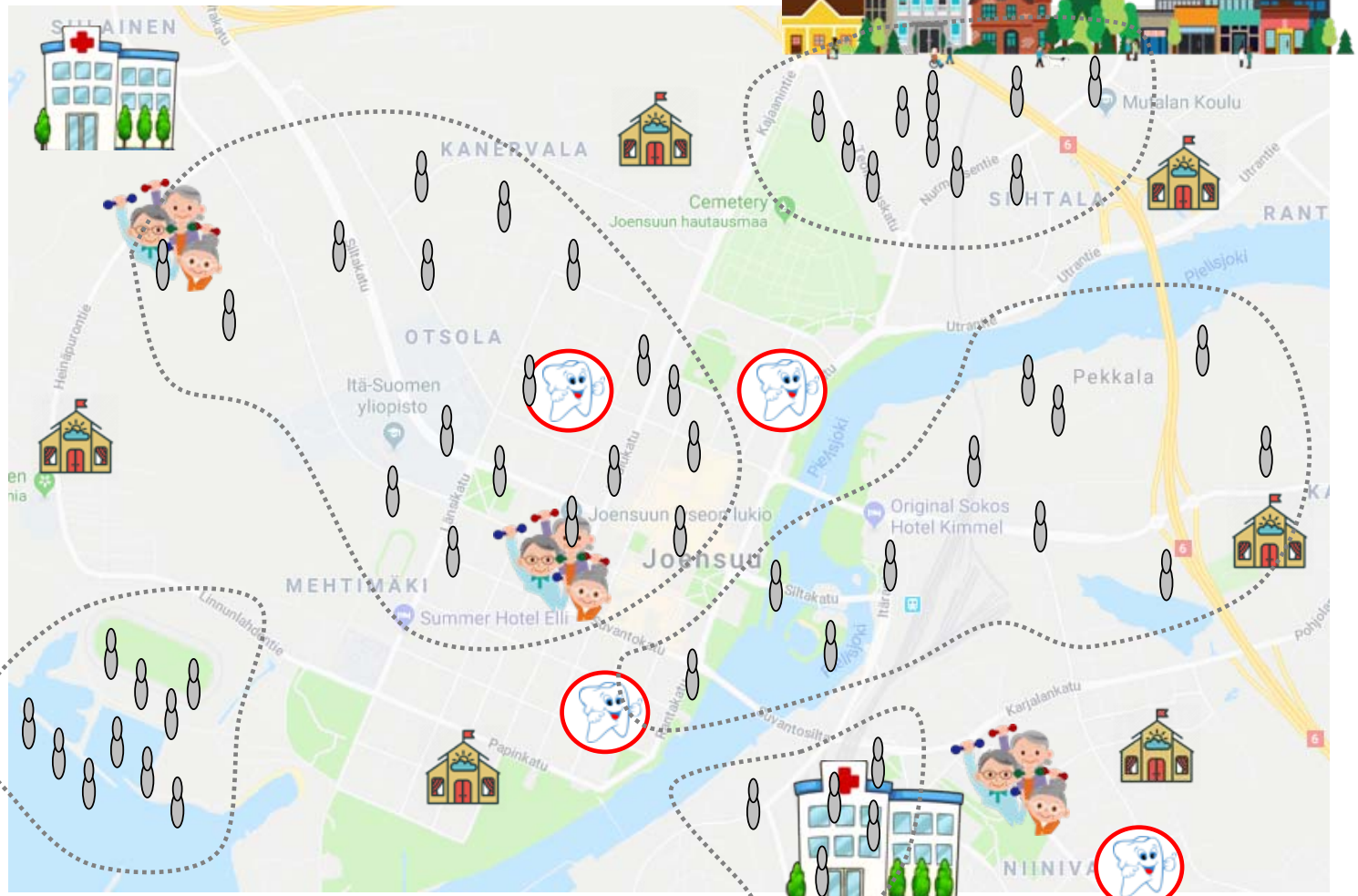
1. Esophageal varices – Fibrosis and Cirrhosis of liver
2. Malignant neoplasm of pancreas – Aftercare
3. Sleep disorders – Non-classified respiratory failure

…

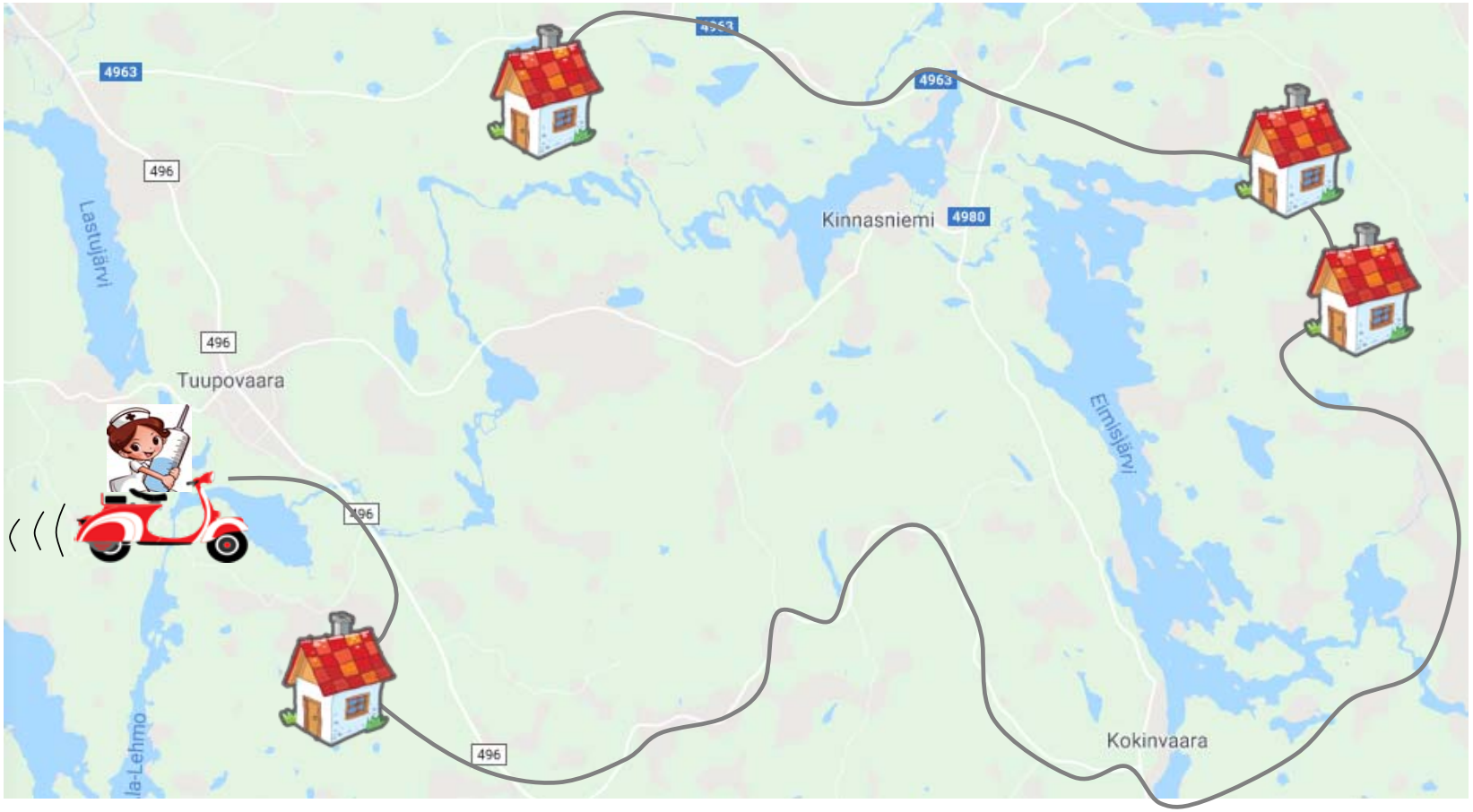
IMPRO

Service locations in city

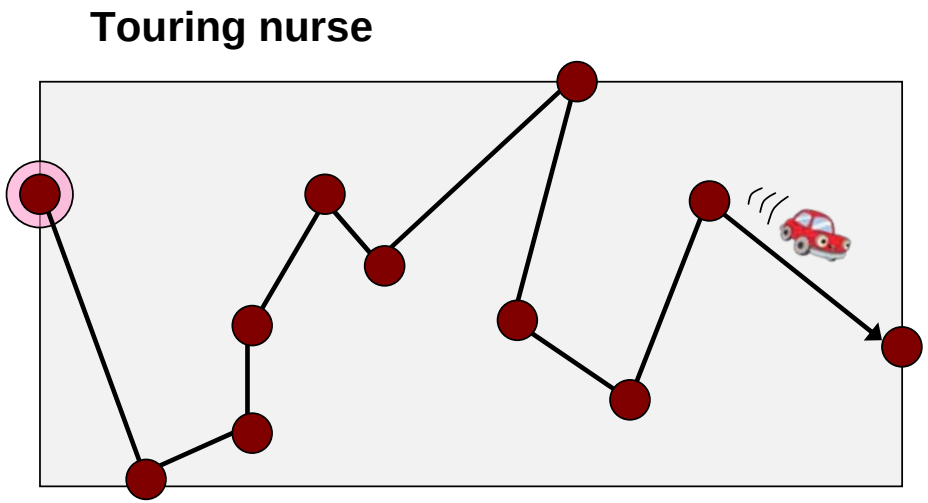
Clustering problem



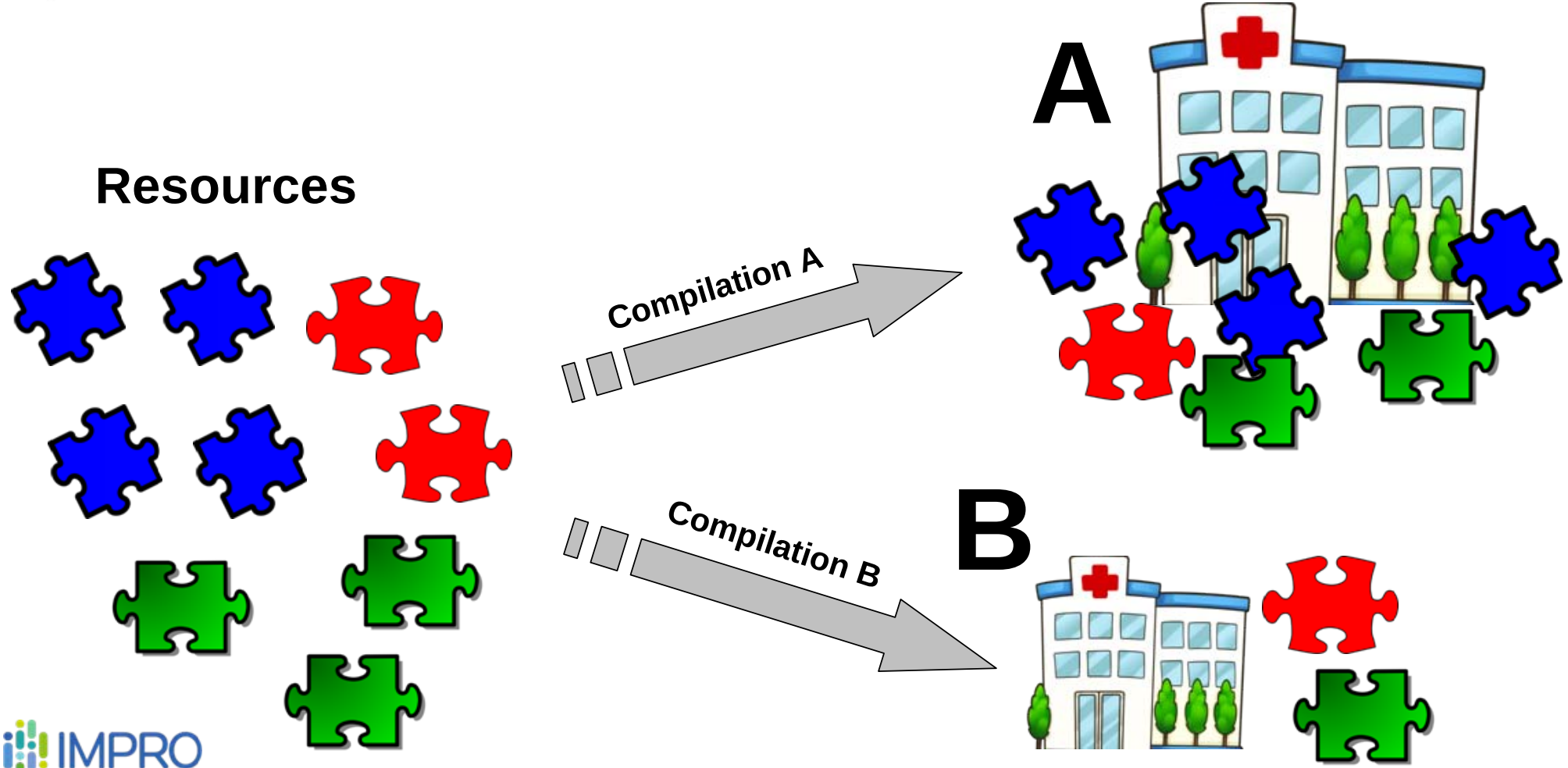
Services in rural area: touring nurse



Travelling salesman problem

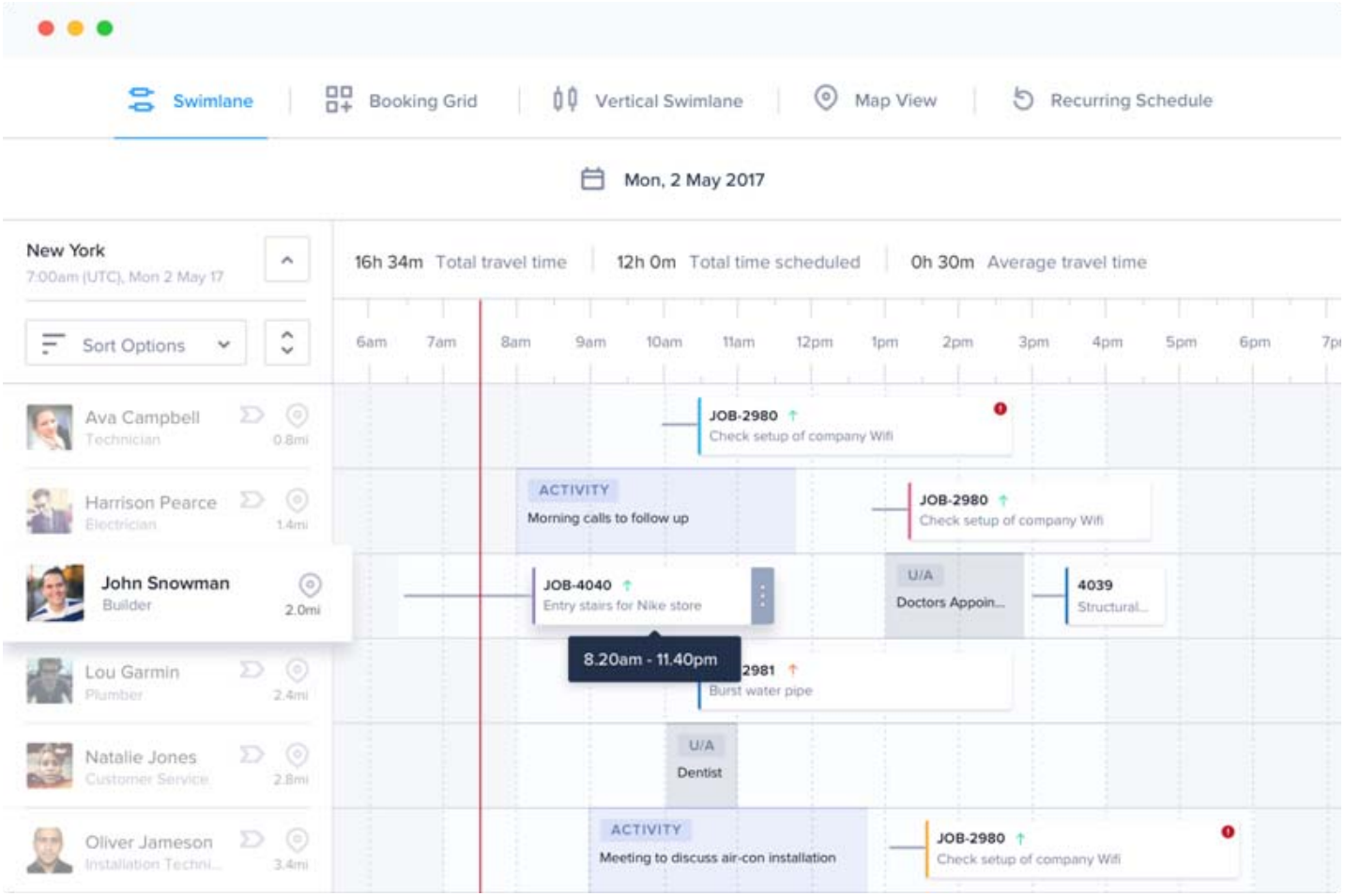


Recompiling the services



IMPRO

Scheduling problem

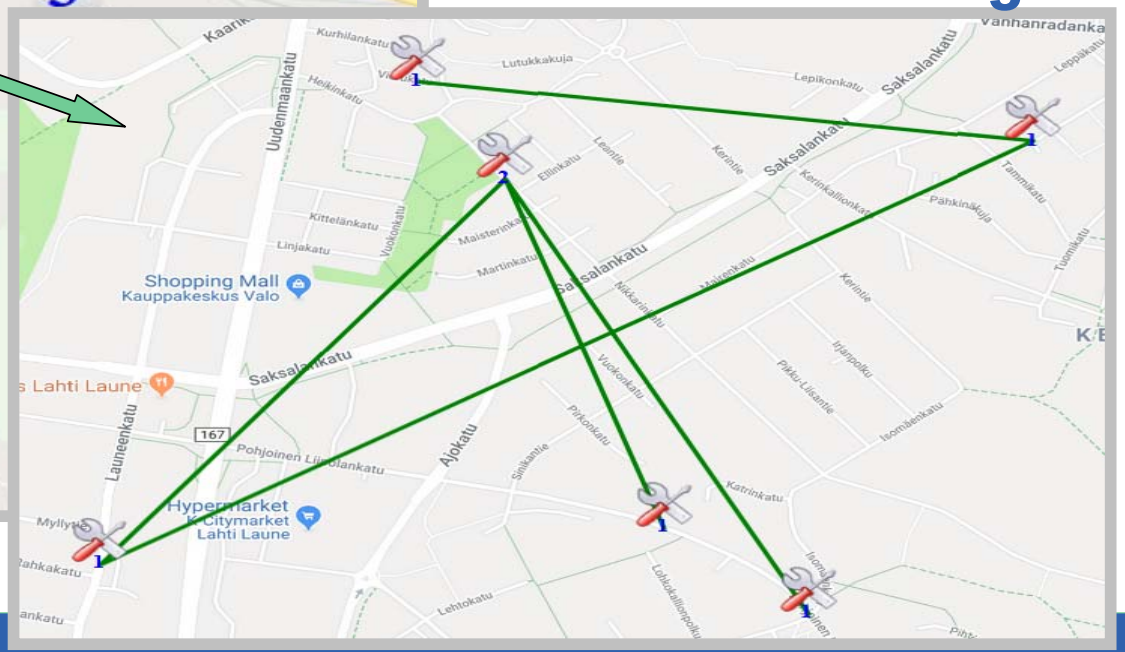


Home take care proto

MOPSI project

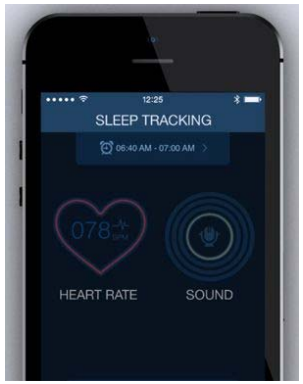
Worker-task allocation

Task scheduling



IMPRO

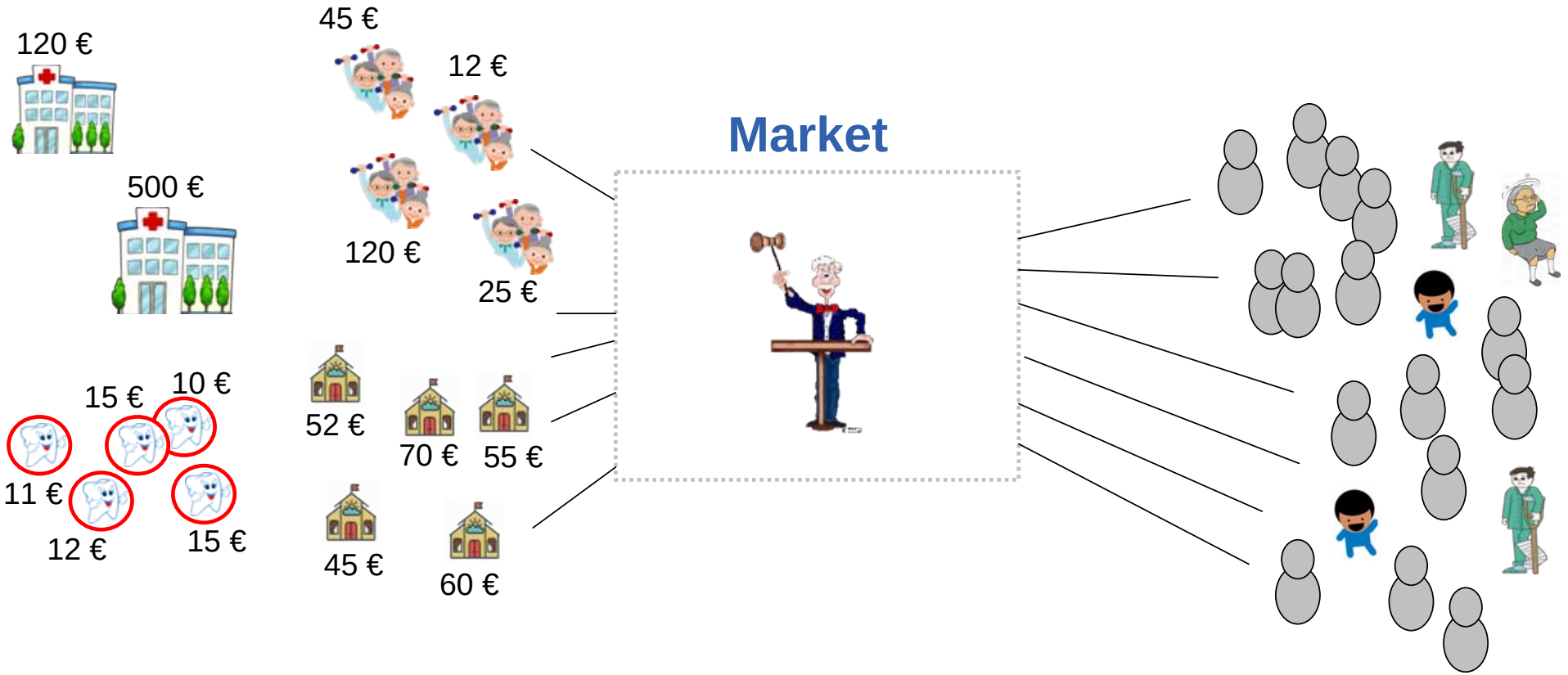
Self-monitoring



IMPRO

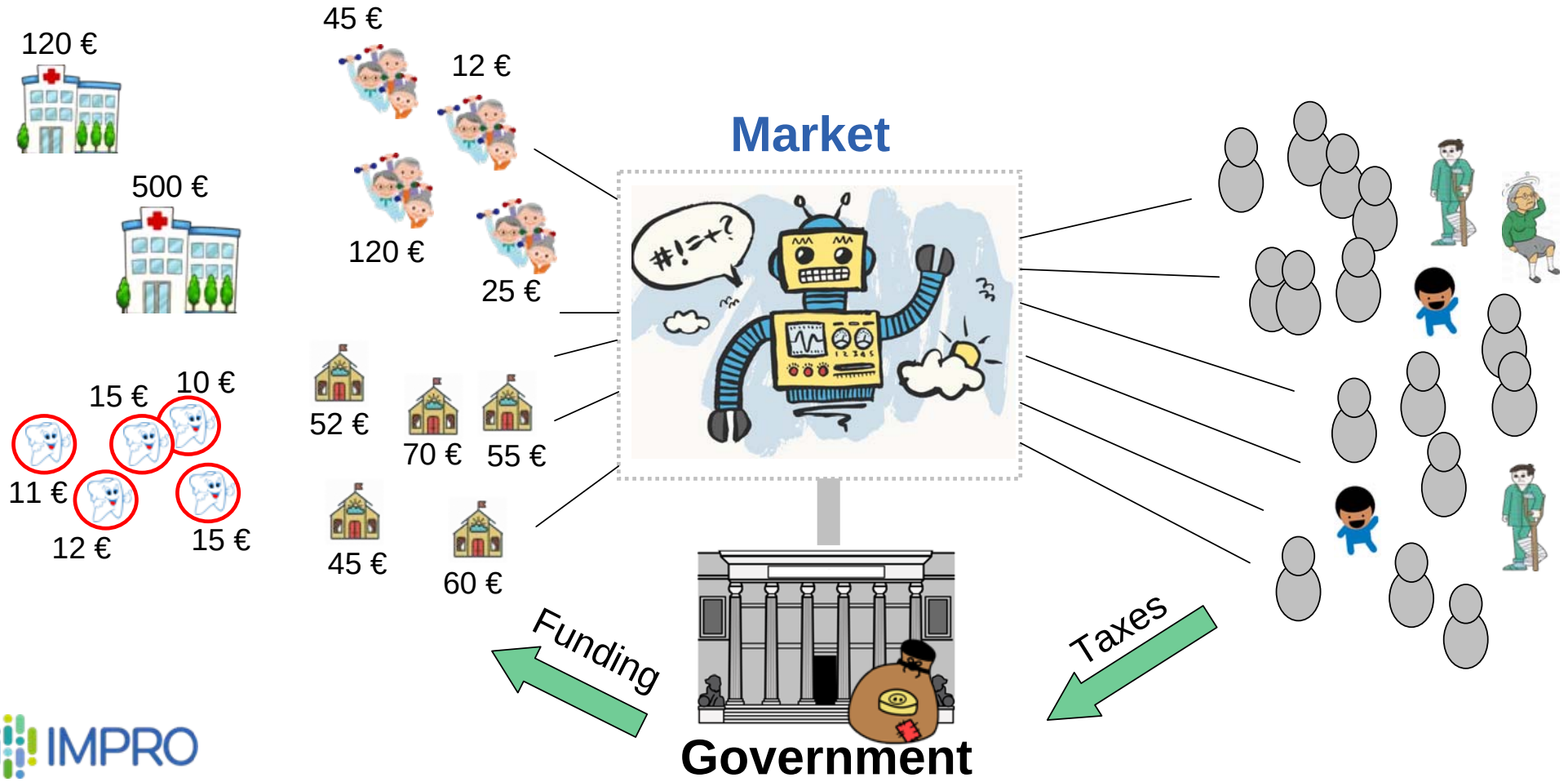
Market place of services

Ideal case



Market place of services

Possible future scenario

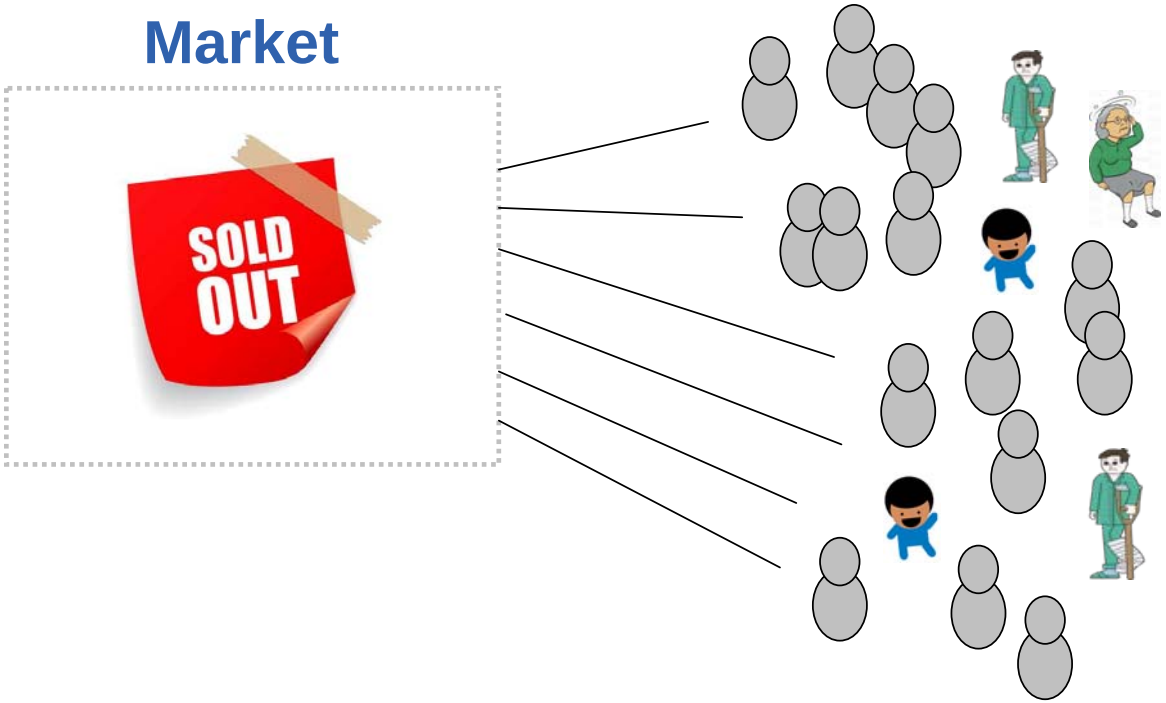


What to optimize exactly?

?

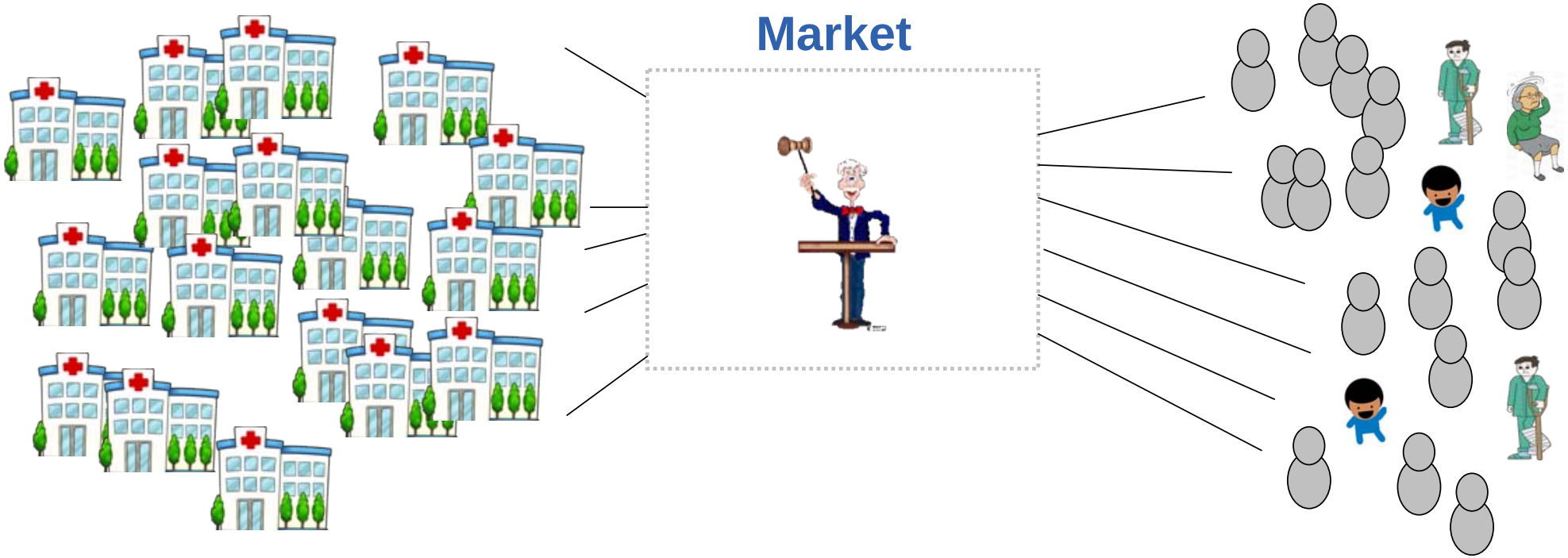
Case 1: No services, no costs

0 €



Case 2: Full service to everyone

1,000,000,000…€



IMPRO

The end

… future under process…

IMPRO

FSDM-2595