

DISCUSSION FORUM DATA AS RESEARCH MATERIAL

CITIZEN MINDSCAPES PROJECT



SPRING 2015: FORMATION OF A RESEARCH COLLECTIVE





PROJECT PARTICIPANTS

Uni Turku

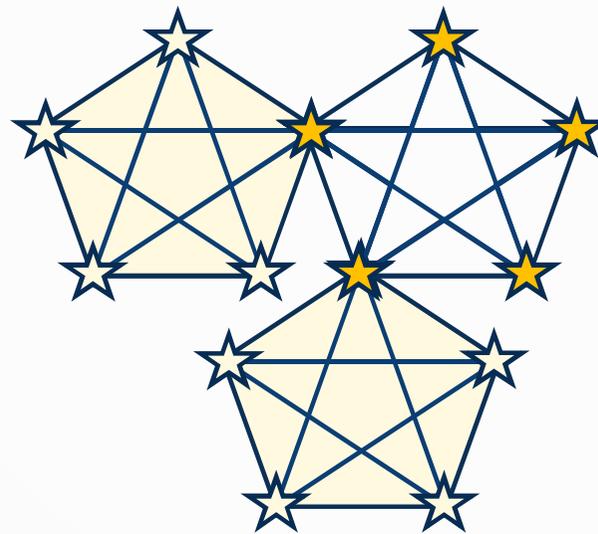
Natural sciences:

Language technology & machine learning methods

Uni Turku

Humanities:

Research ethics
Digital culture of social media



Uni Helsinki

Dept. Political and Economic Research & Consumer Society Research Centre

Social shaping of Suomi24
Waves of emotion & interaction
Connect Lab (larger collective)

Uni Helsinki

Social Statistics:

Statistical modeling of discussion dynamics
Tools

Uni Eastern Finland

Welfare research:

Users, their demographics & wellbeing



CO-OPERATION GOALS

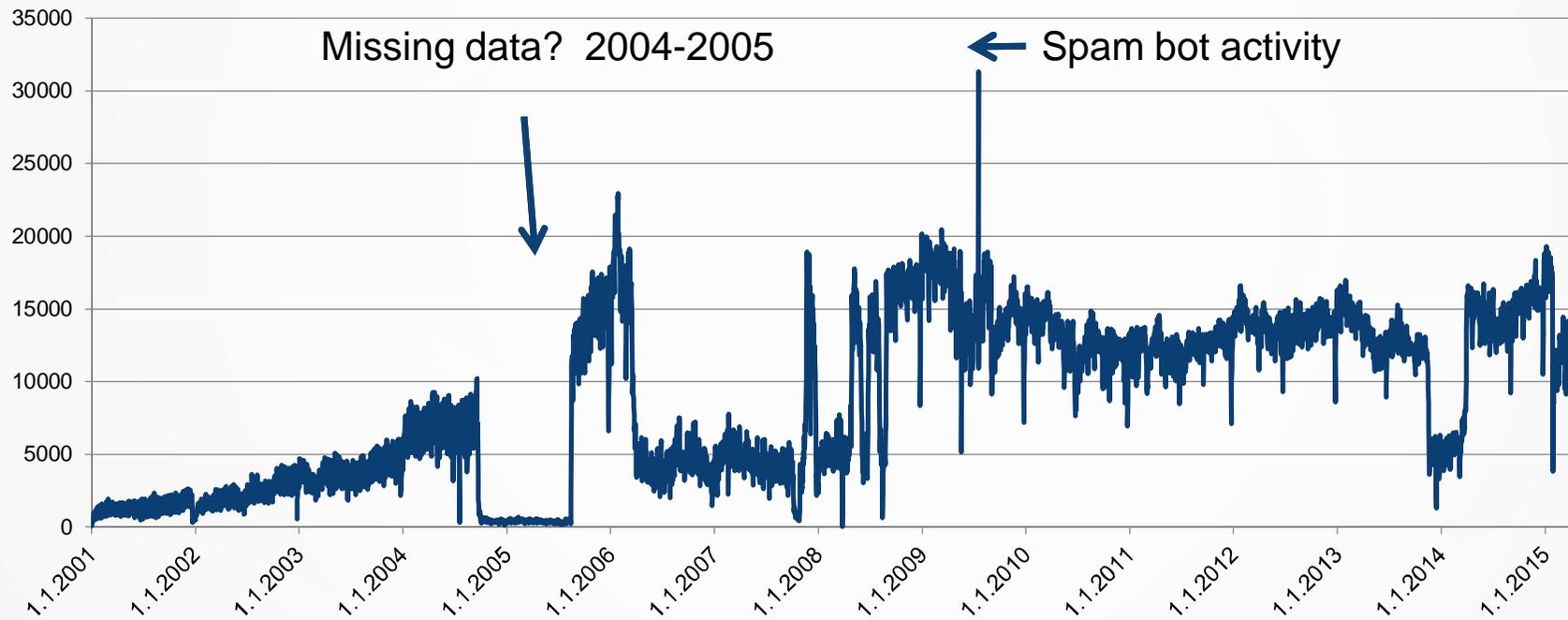
- Widening multidisciplinary network
- Tools for new user groups
 - Korp, social media toolbox
- New skills for researchers
 - Qualitative researchers to use large data sets
- New research problems and questions
- Change in social sciences thinking
 - From individual to team work
- Publications (!)





CORPUS CHARACTERISTICS

- “The corpus contains all the texts from discussion forums of the Suomi24 online social networking website from 2001 to June 2015”
- “The corpus has been tokenized and annotated with morpho-syntactic analysis by FIN-CLARIN at the Department of Modern Languages, University of Helsinki”
- <http://metashare.csc.fi/repository/browse/the-suomi-24-corpora/b4db73da85ce11e4912c005056be118ea699d93902fa49d69b0f4d1e692dd5f1/>
- 231,373,748 sentences
- 48,423,611 texts
- 2,385,073,226 tokens



Publication: "Suomi24 – Muodonantoa aineistolle", Lagus & al, 2016



TOPIC MODELING & SUOMI24 HEALTH DISCUSSIONS

A research case

Keskustelu24 > Ajanviete > Huuhaa > 100 typerystä

100 typerystä

Tee Paita
4.6.2002 23:40

lueke tämän viestin viikon kuluessa.

Jaa | Ilmianna

3 Vastausta KIRJOITA VASTAUS

Äitin poika
5.6.2002 7:00

Vasta ensimmäinen aamu ja jo 15 lukenut viestisi. Mun vanha äitimuori sanoi että kymmenen viisasta ei pysty vastaamaan mitä yksi hullu kysyy. Tämähän ei ollut kysymys, joten nostan hattua loogiselle matematiikalle ja odotan jännityksellä tietoa ensimmäisestä typeryksestä joka viestisi lukee.

Jaa | Ilmianna

LISÄÄ KOMMENTTI

Pöntö
5.6.2002 11:00

Nyt on jo yli 50 typerystä lukenut viestisi. Enää puuttuu puolet!

Jaa | Ilmianna

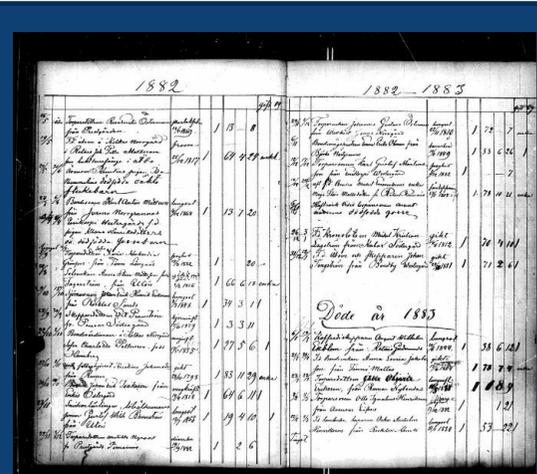
LISÄÄ KOMMENTTI

olen
5.6.2002 19:04

terv.polvi housu

Jaa | Ilmianna

LISÄÄ KOMMENTTI



EKSTIVESTIT

13530

● Karjalainen julkaistiin nimellä Katajalainen. Miellestäni teko oli tilaajien aliarvioimista. Yleisöä oli 2500 ja meitä tilaajia valtavasti, jotka emme seuraa koripalloa.

● Nyt on voitettu etukäteen jääkiekko ja euroviisut, ja vain koripallo oikeasti, et-tä sillee...

● Köyten ja alla?

● Reijlaan p... Pikku... on all...

● On... misia... arma... sukse...

● Vie... ely-k...

CLASSIFIEDS

MORTGAGE SOLUTIONS
We are seeking for legal, ethical, and...
GENERAL MANAGER
Are you looking for a career opportunity...
DANCE CLASSES
Are you looking for a career opportunity...
REAL ESTATE
Are you looking for a career opportunity...
SERVICES
Are you looking for a career opportunity...
EDUCATION
Are you looking for a career opportunity...
NEED A SMALL LOAN
Do you need a small loan for your business...
WEEKEND RECEIPT
Do you need a weekend receipt for your business...
OFFICE AVAILABLE
Do you need an office for your business...
LOOKING FOR WITNESS
Do you need a witness for your business...
GENERAL HELP WANTED
Do you need general help for your business...
GENERAL MANAGER
Are you looking for a career opportunity...
NEED A SMALL LOAN
Do you need a small loan for your business...
WEEKEND RECEIPT
Do you need a weekend receipt for your business...
OFFICE AVAILABLE
Do you need an office for your business...
LOOKING FOR WITNESS
Do you need a witness for your business...
GENERAL HELP WANTED
Do you need general help for your business...

opinion

Editorial

The price is NEVER right

Do you buy all the required textbooks because I need them for my class?

"Yes, I buy all the required books because I need them for my class."

"Yes, I usually go online and buy ebooks to not use my computer or iPad. That way I have less books to carry."

"Yes, I buy my required books in case I need them for class. I try and buy the older editions online to save money."

"Yes, I buy all the required books because they make learning easier. I pay full price because if I don't, I might not get the full material. I don't want my half price and get half a book."

Do you buy all the required textbooks because I need them for my class?



PULSE





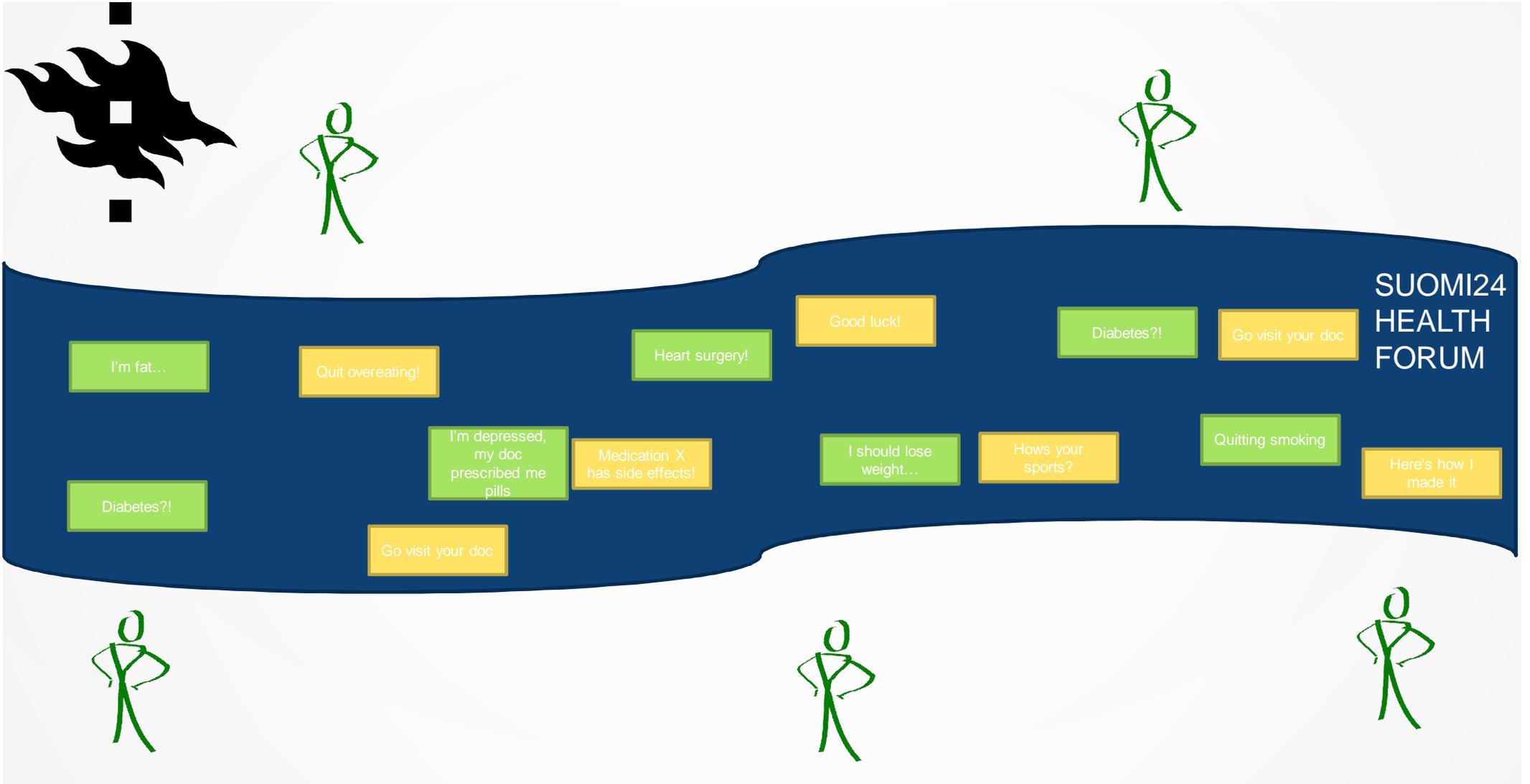
QUESTION: HOW TO DETECT A GROUP PHENOMENON IN BIG TEXT DATA, *THAT RISES AND GOES DOWN IN TIME?*

- Material: Suomi24 online discussion forum data, from 2001 to 2015 (70 million comments) – subforum *Health* (1.5 million comments, ~2.2M pages)
- Case: Low-Carbohydrate Diet phenomenon (LCD)
- Start with static: Latent Dirichlet Allocation (LDA) topic modeling + close reading
- ... but actually, work stopped halfway and something else was found!



CASE LOW-CARB DIET: CONTEXT

- Finland: Fast forward from poor post-war nation, to modern consumer & welfare society
- Diet and health - well-researched topic
 - Prevalence of **diabetes I** in Finland one of the largest in the world; **controllable with low-carb diet**
 - 1970s North Carelia project **suppressed** post-war fat consumption -> today, **backlash?**
 - **Rebellion** against public health nutrition specialists and other authorities?
 - **Food as pleasure**, not fuel or medication?
- Ø Citizens, patients, consumers : roles vary - what is the Unfiltered, Real Voice of The People?
- Ø What diet advice is followed, what ignored? What is the process? Is there a group conclusion when a fad fades? > *Health literacy/health consciousness*





HEALTH FORUM & SUBFORA

keskustelu²⁴

Keskustelu24 > Terveys

Terveys

Keskustele terveydestä ja hyvinvoinnista.

ALOITA UUSI KESKUSTELU

Luetuimpia

- Omien tunteiden ja ajatusten hallinta
Kuuluu ihmiselle itselleen mutta naristit kuvittelee että he hallitsevat sekä omia että muiden tunteita j...
- Paraneeko naristit?
Jos sairastat verenpainetautiä, syöpää, kuumetta tai ilmavaivoja et voi asialle mitään. Voiko naristit its...
- Skitsofrenia ei ole parantumaton
Vaan hetkellinen mielentila. Näin sanoo Pakistinin korkein oikeus. Vaikka kaikki haukkuvat nyt Pakistan...

Keskustelut

- Tukisukien käytöstä kysym...
Niikkani ovat jonkin verran turvoksissa. Hankin Venosan-tukisukat mittojen mukaa...
Tänään 14.04
- Käväisimpäs kaupassa
Käväisimpäs tuossa läheisessä tavaratalossa hieman ostoksilla kun siellä on nyt a...
Tänään 14.02
- Käväisimpäs kaupassa
Käväisimpäs tuossa läheisessä tavaratalossa hieman ostoksilla kun siellä on nyt a...
Tänään 14.02

- Mental health (582,000 comments)
- Illnesses (575k)
- Weight issues (417k)
- Intoxicants (289k)
- Health care (166k)
- Women's health (89k)
- General topics (69k)
- Contraception (62k)
- Medication (61k)
- Dental health (60k)
- Plastic surgery (54k)
- Senses (43k)
- Death and sadness (42k)
- Questions about health (32k)
- Men's health (26k)
- Health care services (2k)



DATA 1/3

- Fetch data from KORP API:

```
https://korp.csc.fi/cgi-bin/korp_download.cgi?format=sentences, csv&subformat=lemmas-  
resultinfo&corpus=S24_009&cqp=[_.text_discussionarea="Terveys.*"]&defaultcontext=1+sentence&start=0&en  
d=50000&show=sentence, lemma&show_struct=sentence_id, text_datefrom, text_timefrom, text_discussionarea, te  
xt_subsections, text_tid, text_cid&structs
```

- Clean duplicates:

```
Import json
```

```
Import re
```

```
From pprint import pprint
```

```
Import datetime
```

```
Import codecs
```

```
Import os
```

```
Impost csv
```



DATA 2/3

hit number	corpus	tokens	lemmas	date	total hits	params
2100041	S24_008	Suosittelen vaihtamaan lekuria , hyvällä tuurilla löydät sellaisen joka ymmärtää asioista jotakin , tuo nykyinen lääkärisi vaikuttaa huonolta ja saat kärvistellä vaivojesi kanssa liian kauan .	suositella vaihtaa lekuria , hyvä tuuri löytää sellainen joka ymmärtää asia jokin , tuo nykyinen lääkäri vaikuttaa huono ja saada kärvistellä vaiva kanssa liian kauan .	#####	12692668	end=2125000 corpus=S24_008; cqp=[_.text_discussionare a="Terveys.*"]; defaultcontext=1 sentence; defaultwithin=[none]; sort=none; start=2100001;



DATA 3/3

```
import logging
import os
from gensim import corpora, models,
similarities, utils
from gensim.corpora import TextCorpus,
MmCorpus, Dictionary
from collections import defaultdict
from pprint import pprint
import json
import csv
import nltk
import nltk.data
from nltk import snowball
from nltk.corpus import stopwords
```

- Omit
 - Finnish stopwords
 - Words appearing less than 10 times
 - `/p p a quot > < () ! , . : ; & ? * NUOLI + [] ... / # ' ' -- -blank /a " " http paypal`

Ø What remains is colloquialisms like *mä* (“me”)



RESULTS

6 – 8 – 20 – 30 – 50 topics using subsets and full sets

MEANING OF LIFE:
Life, world, other, human

INTOXICANTS:
Alcohol, tobacco,
cannabis, drug, use

MENTAL HEALTH:
Depression, suicide,
psychiatrist,
therapist, feeling,
bad, tolerate,
anymore, worry,

WEIGHT ISSUES:
Weight, fat, thin,
overweight, food,
eat, dieting

OTHER:
Operation, eye,
heart, cancer, pain,
narcissism, stress,
...

SUSPECT

GET+INFO

PATIENT

TREAT

MEDICATION

TOTALLY
SURVIVE

SYMPTOM

FIND

NARRATIVE OF DISEASE:

WRITE

POSSIBLE

WAIT

PRESCRIBE

HELP

RELIEVE

FALL+ILL

FEAR

MEDICAL+EXAMINATION+RESULT

REGIMEN

HEAL

AUTHORITY-
ADVICE:
Treatment,
better, visit,
physician,
receive, help,
take

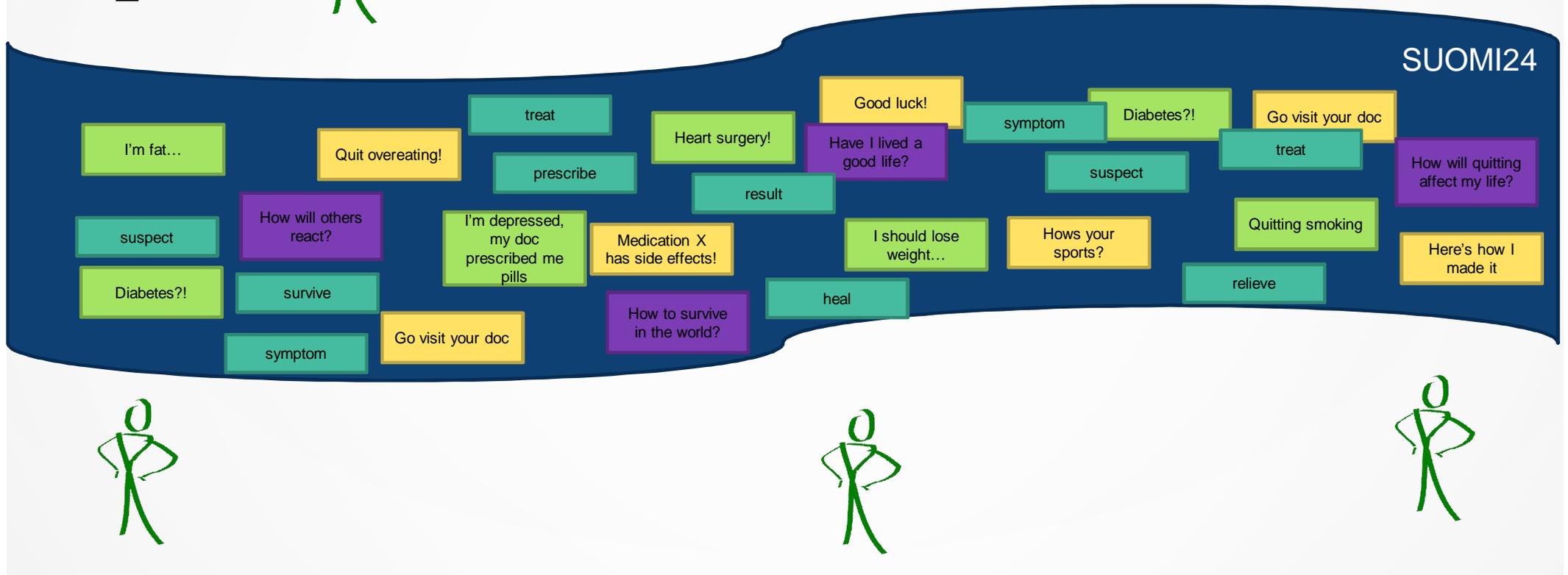
EMOTIONAL
ADVICE:
"TSEMPPIÄ"
~ Good luck!

EXPERIENCE-
ADVICE:
Times, day,
week, month,
year, after, next

MONEY:
Price, cost,
expensive

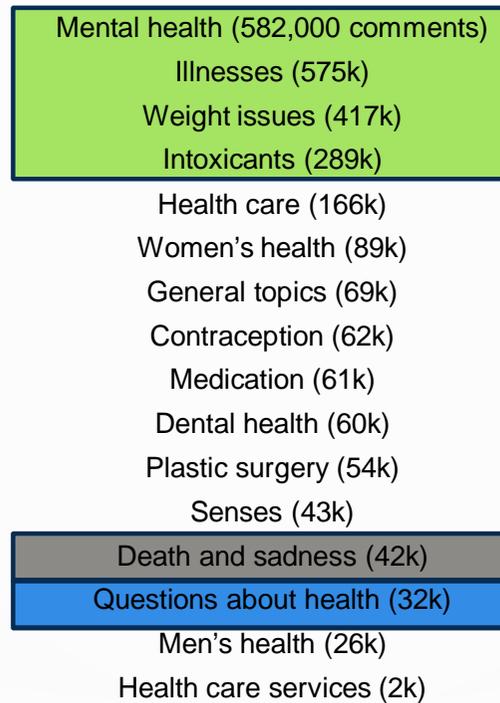


INTERIM CONCLUSIONS



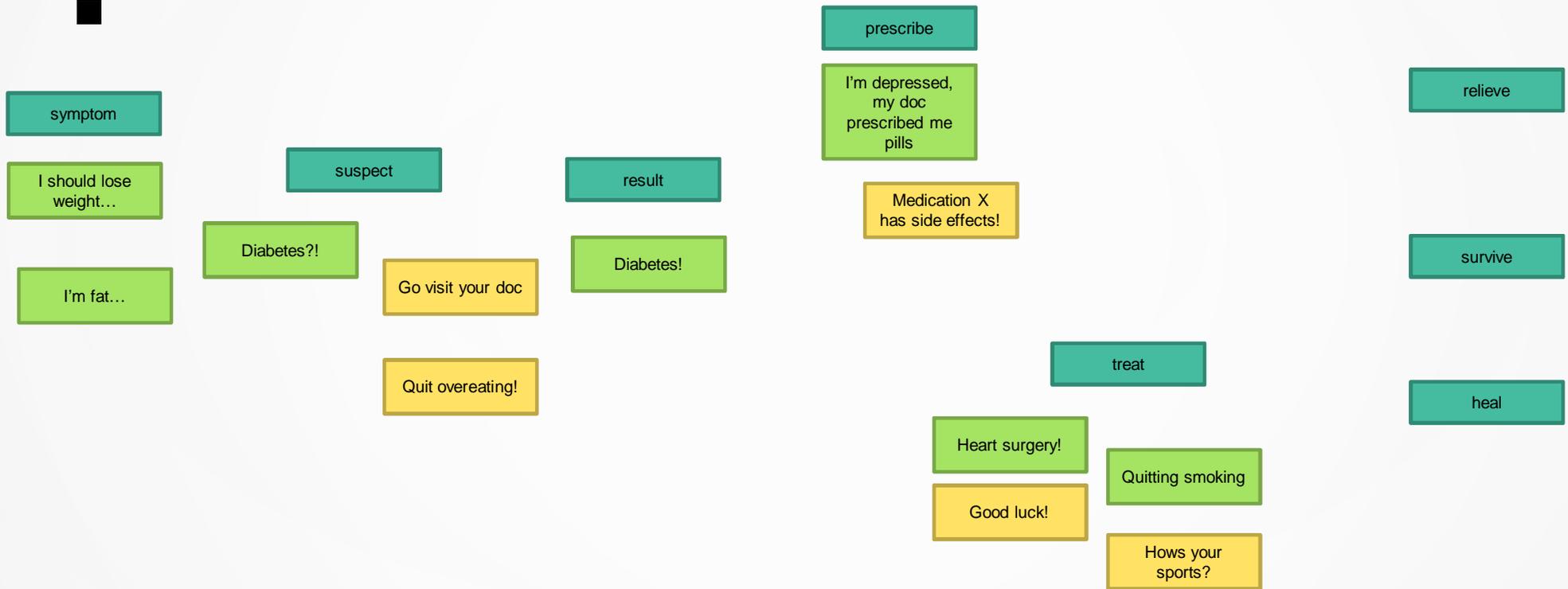


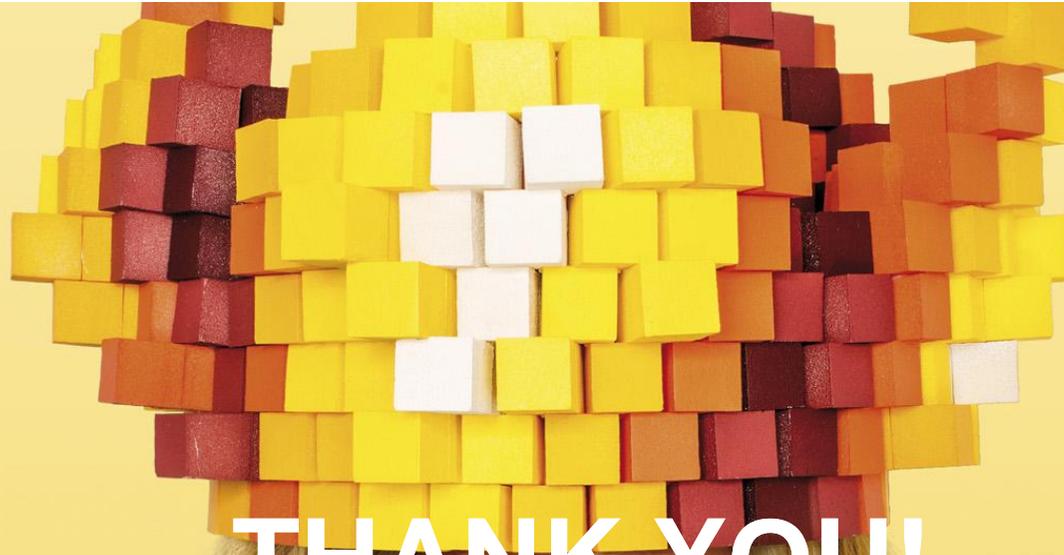
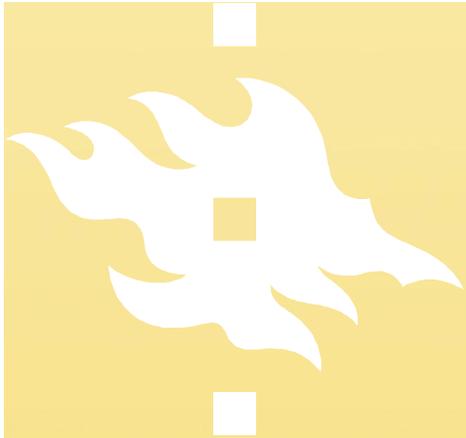
HEALTH FORUM SUBAREAS





GENERAL NARRATIVE OF ILLNESS IN CITIZEN CONVERSATIONS ON HEALTH





THANK YOU!

Marjoriikka.ylisiurua@helsinki.fi

**HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI**