

# Loppuraportointi

## Data-analytiikka ja paikkatieto - 2021

### Kaupunkilaisten kehitysideat kartalla

Helsinki



Tukea digitalisaatiokokeiluihin  
kaupungin työntekijöille

# Kaupunkilaisten kehitysideat kartalla

Kaisa Koskinen, Tiina Saukkonen, Sari Yli-Säntti, Riku  
Ruuskanen, Anna Böhling

Riku Oja ja Santtu Pyykkönen, Gispo Oy

# Kokeilun yhteenvedo

Pystymmekö automaattisesti luokittelemaan kaupunkilaisten kehitysehdotukset ja esittämään ne kartalla lämpökarttana?

## Anna palautetta

Anna palautetta tai kysy kaupungilta. Pyrimme vastaamaan viiden arkipäivän kuluessa. Jos asia edellyttää selvitystyötä tai suunnittelua, vastaaminen voi kestää kauemmin.

Voit lähettää viestisi nimettömänä. Jos haluat henkilökohtaisen vastauksen, jätä yhteystietosi.

Palautteesi voidaan julkaista luvallasi.

### Julkaistut palaukset

#### Palautteen tyyppi

- Kiitos
- Moite
- Kysymys
- Kehitysehdotus
- Verkkopalvelujen saavutettavuus
- Muu

#### Otsikko

#### Viesti (pakollinen)

(enintään 10000 merkkiä)

- Palautteeni saa julkaista kaupungin verkkosivuilla
- Palautettani ei saa julkaista kaupungin verkkosivuilla

#### Palautteeni liittyy tähän paikkaan

TAI

#### Liitetiedosto (max. 5 Mb)

No file chosen

No file chosen

No file chosen

laivosta, emme kaipaa pyöreäsi teräviä hiekkosuosseja punnimaan teräksilläme, meina on nastarenkkaat polkupyörissämme. Vai haluaako Helsinki kannustaa myös työmatkapyöräilijät yksityisautolleen? Ja edelleen kaduille ei nyt ole hiekoittamista kaipaavaa jättä, siellä on LUNTA. Kertä varten tätä kaupunkia oikein tehdään?

#### Vastaus

10.01.2022 15:01 - Kaupunkiympäristö  
Hel.

Kiitos viestistänne. Lähtökohhta on sa, ette ne viylät mahdollisimman turvalliseen kunnessa ja kadunpitiäänä pyrimme huolehtimaan tästä asiasta. Jos hiekoitamme vain puoliaki jalkakävelyä, niin tämä voi aiheuttaa kaatumisia ja korvauksia kaupunkille. Työnjohto seuraa sääitä ja suorittaa liukkaudentorjuntaa sääolosuhteiden mukaan.

Ystävällisin terveisin

Kaupunkiympäristön asiakaspalvelu  
PL 58231, 00099 Helsingin kaupunki  
Työpajankatu 8  
p. +358931022111  
puhelinpalvelu ma - to 8-15 - 16, pe 12 - 15  
[www.hel.fi/kaupunkiymparisto](http://www.hel.fi/kaupunkiymparisto)

#### Sijainti

Koordinaatit: (25498415, 6677100)



#### Palautteen käsittely

07.01.2022 10:37 - Saapunut



# Kokeilun yhteenveto

- Palautteen tyyppiä tai otsikkoa ei saada rajapinnasta
- Onnistuimme ryhmittelemään palautteet automaattisesti luokkiin (jokaisella palautteella yhdestä useampaan luokkaa):
  - Vaara, Luonto, Vesi, Liikunta, Valaistus, Opasteet, Kalusteet, Kiitos, Jalankulku, Pyöräliikenne, Joukkoliikenne, Autoliikenne, Pysäköinti, Palvelut, Koira, Kunnossapito, Roskat, Työmaat
  - 5 % palautteista ei saanut yhtäkään luokkaa
    - palautteen antaja oli kirjoittanut asiansa lomakkeen Otsikko-kenttään ja Viesti-kenttään kirjoittanut esim. "katso yllä"
    - Vieraskielinen palaute
    - Palaute annettu kuvana, tekstissä esim. "Katso kuva" tai "Tällaista täällä"

# Tekstianalyysi

- Perustui suomen kielen sanojen perusmuotoistamiseen, jäsentimenä suomalainen avoin  
<https://github.com/EvidentSolutions/elasticsearchanalysis-raudikko>
- 1) Laskettiin palautteissa esiintyvät perusmuotoiset sanat esiintyvyyssjärjestyksessä =>
  - 2) Poistettiin analyysistä merkityksettömät/sisällöttömät/turhat sanat =>
  - 3) Lähdettiin listaamaan esiintyvyyssjärjestyksessä yleisimpiä avainsanoja sen mukaan, minkätyyppiseen palautteeseen mikäkin viittaa. Iteratiivisesti käytiin läpi vielä luokittelemattomia palautteita puuttuvien avainsanojen lisäämiseksi.

Tällöin jokaiselle palautteelle tulee kaikki ne kategoriat, joihin viittaavia sanoja palautteessa esiintyy.

# Palautteet kartalle

- Suuri osa palautteista sisälsi koordinaatit valmiina
- Noin tuhat palautetta sisälsi paikannimiä, mutta ei koordinaatteja.
  - Palautteet, joissa oli paikannimiä, perusmuotoistettiin, ja perusmuotoisen paikannimen perusteella palautteelle lisättiin koordinaatit Helsingin nimistö -WFS-aineiston perusteella.
- Lopputuotteena QGIS-projekti, jossa palautteet kartalla kategorioittain eri tasoilla pisteinä, käyttäjän valittavissa mitä kategorioita katsellaan. Lisäksi lämpökartta, jossa katseltavissa tihentyminä yksi kategoria kerrallaan.

# 1. Kokeilun onnistuminen

- Kokeilu osoitti, että yhdistämällä varsin yksinkertainen tekstianalyysi karttatietoon on mahdollista tuottaa yllättävän paljon arvoa KAMU-palvelussa työskenteleville asiantuntijoille.
- Ketterä kehittäminen menetelmänä vaatii uuden tyyppistä työn organisointia kaupungilla. Tämä ei onnistunut parhaalla mahdollisella tavalla tiimissämme. Paljon asiantuntemusta jäi hyödyntämättä tämän takia.

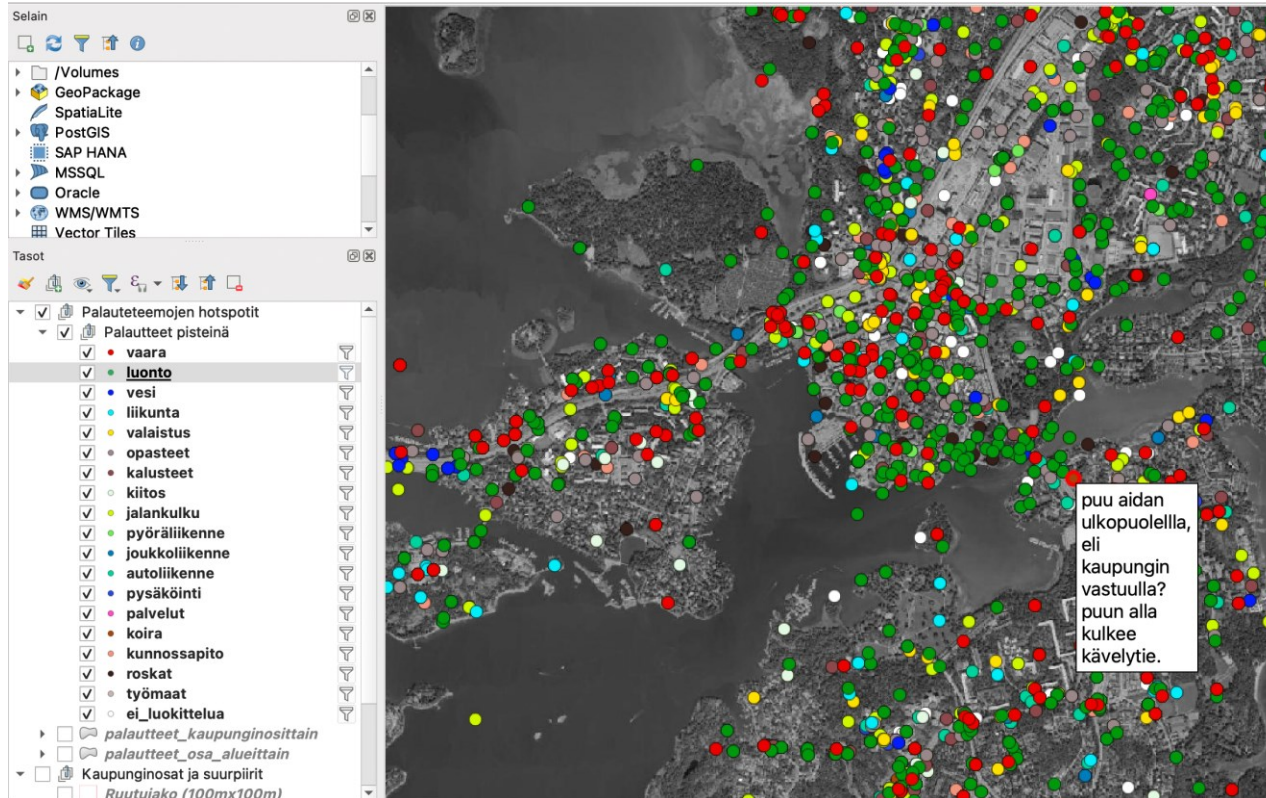
## 2. Kokeilun eteneminen

1. Pohjadataa käytettiin HRI:n avoimesta rajapinnasta saadut asiakaspalautteet. Avoimesta rajapinnasta saa vain 200 palautetta kerrallaan ja vain yhden vuoden ajalta.
2. Prototyyppi datan prosessoinnista, tekstianalyysistä ja sijaintitiedon hyödyntämisestä tehtiin vuoden 2020 datalla Github-repoon.
3. Tekstianalyysiä ohjattiin ja hiottiin yhdessä projektitiimissä substanssiasiantuntijoiden voimin.
4. Tekstianalyysin tuloksilla jalostetut palautteiden sijainnit vietiin QGIS-projektiksi
5. Vuoden 2020 datasta skaalattiin analyysi neljän vuoden data-aineistolla.
6. Datan visualisaatioiden iteraatioiden avulla tuotettiin substanssiasiantuntijoiden näkökulmasta hyödyllisiä näkymiä dataan.



# 3. Kokeilun tuotokset

- Avoimena: <https://github.com/gispoCoding/geospatial-text-analysis>



# 4. Opit data-analytiikan ja/tai paikkatietojen mahdollisuuksista

- Lähes kaikki kaupunkilaisen palaute on sijaintipohjaista eli kaupunkilaisen positiiviset ja negatiiviset palautteet liittyvät lähes aina kohdennetusti johonkin tiettyyn sijaintiin.
- Kokeilussa keskityttiin yleisten alueiden suunnitteluprosessiin, mutta luotu ratkaisu on skaalattavissa muille yksiköille. Tärkeätä on saada data karttapohjaisesti näkyviin eri substanssiasiantuntijoille (esim. talvikunnossapito, kevyen liikenteen väylät jne.)
- Kaupunkilaisten palautetietojen yhdistäminen muihin kaupungin paikkatietoaineistoihin antaa substanssiasiantuntijalle paljon lisäarvoa (esim. Penkkien sijainti, alueen ikäjakauma ja esteettömyyden tavoitetaso)

# 5. Opit data-analytiikka- ja/tai paikkatietoratkaisujen kehittämisestä

- Suomen kielen jäsennys on hyvä työkalu, mutta vain yksittäisiä sanoja luettelemalla ei koskaan päästä parhaaseen tarkkuuteen
- Parempi tekstin teeman tulkinta ja havaitseminen edellyttäisi isompa tekoälyprojektia
- Paikkatiedot olivat datassa suurimmalta osin kunnossa, ja puuttuvat tiedot voi hyvin lisätä tekstin paikannimien perusteella

# 6. Opit kokeilemisesta

- Avoin dokumentaati ja sovelluskehityksen prosessit Github-alustaa hyödyntäen toimi hyvin. Mahdollistaa skaalaamisen ja jatkokehittämisen, eikä aiheuta toimittajalukkoja tilaajalle.
  - Tässä kokeilussa ei käytetty kaupungin tekoälyprojekti-github-mallia, mutta ratkaisun kehittämisessä pohjauduttiin samaan periaatteeseen.
- Missä vaiheessa pitää lopettaa? On aina haastava lopettaa, kun on päästy vauhtiin kehittämisessä ja kokeilussa. On hyvä muistaa kiihdyttämön perusajatusta noudattaen, että keskeneräinenkin kokeilu opettaa jo paljon.

# 7. Opit resursoinnista

- Suuri osa ajasta meni käytännön datankäsittelyyn, visualisointiin, datan katseluun ja luokitteluun sekä skriptien kirjoittamiseen
- Budjetti riitti juuri sopivasti kokeilun tarpeisiin, eli perustoteutukseen
  - Iteraatiokierroksia, luokittelun hiomista ja visualisaation hiomista olisi toki voinut olla useampia
  - Luokittelupalaverit kaupungin työntekijöiden kanssa olisivat ehkä auttaneet enemmän luokittelun hiomisessa
- Myös kaupungin työntekijöillä tulisi siis olla aikaa osallistua käytännön toteutukseen, niin yksityiskohdista tulisi parempia.
- Virkahenkilöiden käytetty työaika luovaan pohdiskeluun on tärkeätä merkittävien etujen saavuttamiseksi. Teknologia voi mahdollistaa huimia asioita, mutta vain kiitos hyvin jäsennellyn ongelman.

# 8. Kokeilun tekninen ympäristö

- Avoin lähdekoodi, avoin data => avoin lopputulos
  - 1) Lopputuotteeseen eli tuloksiin + QGIS-projektiin on linkit suoraan projektin [Wikistä](#), tulokset koodin yhteydessä githubissa
  - 2) Analyysitoteutus ladattavissa githubista ja käytettävissä millä tahansa koneella, tarvitaan koodausosaamista mutta ei pilvipalvelualustaa:
    - <https://github.com/gispoCoding/geospatial-text-analysis>
    - Teknisenä ratkaisuna Python + Jupyter Notebook
    - Analyysiin tarvitaan Elasticsearch + jäsennin, asennus Docker-ympäristöön linkistä löytyvillä ohjeilla
    - Olisi hyvä, että joku ulkopuolinen taho/kehittäjä kokeilisi, saako ympäristön pystyyn ja analyysin ajettua omin avuin? Eli ovatko ohjeet ja dokumentaatio riittävät vai puuttuuko jotakin?

# 9. Kokeilun data

- Datana ensisijaisesti HRI:ssä listattu avoin data.  
Palauterajapinta <https://dev.hel.fi/apis/open311>
- Ongelmallista avoimessa datassa oli, että palautedataa **ei ollutkaan avoimena** useammalta vuodelta
  - Avoimesta rajapinnasta on ladattavissa vain viimeisen vuoden palautteet, vanhempia ei löydy
- Suurempi datamassa (neljän vuoden avoin data) jouduttiin pyytämään erikseen tietokantadumppina, eli data ei ollutkaan avointa
- Palautteiden kategoriat jouduttiin pyytämään erikseen tietokantadumppina, niitä ei ollut saatavilla muualla
- Opit: avoin data ei aina ole oikeasti avointa!
- Käyttäjän pitää erikseen valita, että hänen vastauksensa saa julkaista. Kuitenkin hyvin harva palautteista on sellaisia, että niistä kävisi ilmi yksilöllisiä tietoja. Olisi syytä pohtia, kuinka palautteet voisi anonymisoida ja julkaista automaattisesti.

# 10. Jatkopäätökset ja -ideat

- Jatkamme tekstin luokittelua nostamalla penkkeihin liittyvät asiakaspalautteet omaksi luokaksi.
- Jatkamme projektia koko aineistolla. Nyt käytössä vain 10 % aineistosta.
- Tuotantokelpoisen ratkaisun toteuttamiseksi olisi hyvä iteroida temaattisia tekstianalyysiluokkia (esim. Kunnossapito, luonto) ja niihin liittyviä sanoja yhdessä substanssiasiantuntijoiden kanssa.
- Tuotantokelpoinen ratkaisu sellaisenaan ilman jatkokehittämistä ei ole pitkän matkan takana. Mikäli palauterajapinta tuottaa saman tietorakenteen mukaista dataa, voi tekstianalyysin aina ajaa uudestaan uudella datalla.