

Tekoälykokeilun loppuraportti

Tekstianalytiikka avuksi
selvittämään: Miten viestimme
kaupunkistrategian toteutumisesta?

Maaliskuu 2020

Helsinki



Tukea digitalisaatiokokeiluihin
kaupungin työntekijöille

Tekstianalytiikka avuksi selvittämään: Miten viestimme kaupunkistrategian toteutumisesta?

Jani Merimaa

Alexi Salonen

Ville Henttonen / Helsinki Intelligence

Olli Parviainen / Sometrik

Maaliskuu 2020

Helsinki

1. Kokeilun onnistuminen

1. Lopputuotos

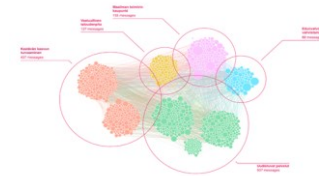
- Saimme aikaiseksi lopputuotoksen
- Lopputuotos kertoo suuntaa-antavan tuloksen
 - Uskottava suuruusjärjestys

2. Tekoäly ja koneoppiminen

- Saimme oppia mallin luomisesta ja koneen opettamisesta
- Ymmärrys tekstianalytiikan käyttötavoista kasvoi kokeilun yhteydessä

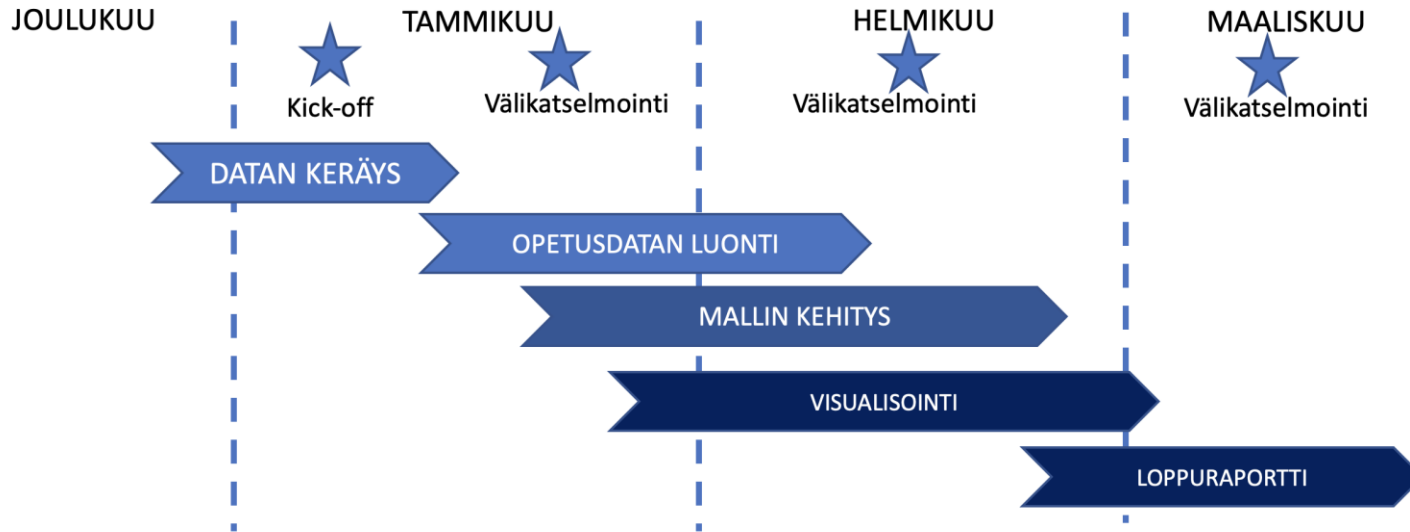
3. Tekoälyn soveltaminen

- Kokeilu antoi ajatuksia uusista soveltamisalueista



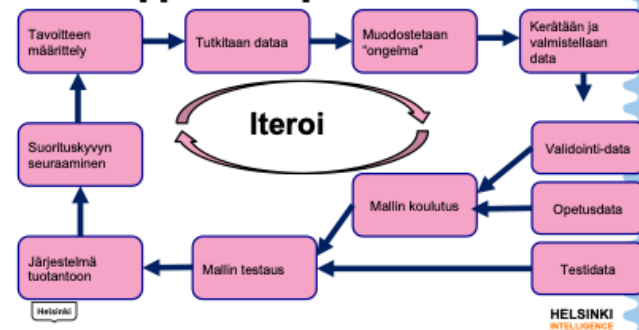
2. Kokeilun eteneminen

- Kokeilu eteni alkuperäisen suunnitelman mukaisesti



Opetusdatasta tekstinluokittelumalliksi

Koneoppimisen prosessi



Opetusdatan syöttö



Opetusdata



Mallin opetus



Paragraph in Helsingin kaupunki

Helsingin kaupungin ICT-kehittämisenetelmit auttavat rakentamaan tietoturvan ja tietosuojaan osaksi palveluita. Kehittämisenetelmiä seuraamalla tietoturva ja tietosuoja tulee huomioiduksi koko järjestelmän elinkaaren ajan.

Helsingin kaupunki | unknown at 9.12.2019 10:16:15 | #7742

- ☐ **Toimivuus** Maailman toimivin kaupunki
- ☐ **Kestävä kasvu** Kestävän kasvun turvaaminen
- ☐ **Palvelut** Uudistuvat palvelut
- ☐ **Taloudellisuus** Vastuullinen taloudenpito
- ☐ **Edunvalvonta** Edunvalvonta | Yhteistyö
- ☐ **Ei strategiaa** Ei strategiaa

Send Skip

Helsinki

Arviointia yhteensä: 1740

Maailman toimivin kaupunki	108
Kestävän kehityksen turvaaminen	68
Uudistuvat palvelut	155
Vastuullinen taloudenpito	28
Edunvalvonta ja yhteistyö	52
Kiinteistöstrategia	16
Kävelykeskusta ja maanalaisten kokoojakatu	4
Maapolitiikka	28
Merellinen strategia	87
Eriarvoisuuden vähentämisen ja nuorten syrjäytymisen ehkäisy	44
Liikkuishuolto	127
Hyvinvoinnin ja terveyden edistäminen	21
Ei strategiaa	1088

SOMETRIK

fastText



Facebook
Open Source

HELSINKI
INTELLIGENCE

Koneoppimismallin toimivuuden validointi

text	strategy	FUNCTIONALITY	SUSTAINABILITY	SERVICES	MANAGEMENT	LOBBYING	NONE
Töölönlahden liikkijat ovat saaneet maailmanympärysmatkan päätökseen. Liikkumishaasteen 40 000 kilometriä tul	SERVICES	0 %	14 %	85 %	0 %	1 %	0 %
Haluamme olla maailman toimivin ja parhaiten digitalisaatiota hyödyntävä kaupunki maailmassa 🤖 Miten tämä oikei	SERVICES	49 %	0 %	51 %	0 %	0 %	0 %
Parempaa valaistusta rantareiteille? Osallistu Helsingin kaupunkiympäristö n työpajaan 🛠️🔧🔦	FUNCTIONALITY	90 %	1 %	9 %	1 %	0 %	0 %
Messukeskus on saanut katolleen uuden aurinkovoimalan. Helen in 1600 aurinkopaneelin vuosituotto on aluksi n	SERVICES	0 %	5 %	95 %	0 %	0 %	0 %
Joko olet käynyt retkellä Vasikkasaarella? Tänä kesänä saari on saanut kaupungin uuden ilmeen mukaiset opaste	SERVICES	0 %	0 %	100 %	0 %	0 %	0 %
🌊 Lähtisitkö kesäiselle meriretkelle? Tänä kesänä retken suunnittelu on aiempaa helpompaa, sillä HSL :n Reittio	SERVICES	0 %	0 %	100 %	0 %	0 %	0 %
Kaupunkiveneet rantautuvat Helsinkiin toukokuussa! Ensimmäiset venepisteet tulevat Töölönlahdelle, Munkkinieme	SERVICES	0 %	0 %	100 %	0 %	0 %	0 %
👩🏻👧🏻 Hurraa! Helsingin kaupunkipyöriä käytetään ahkerammin kuin missään muualla maailmassa 🚲👩🏻👧🏻	SERVICES	3 %	9 %	87 %	0 %	0 %	0 %
👩🏻👧🏻 Kahden viikon kuluttua poljetaan 🚲👩🏻👧🏻 Kaupunkipyöräkausi starttaa 1.4. ja toukokuussa palvelu laajenee H	SERVICES	0 %	3 %	96 %	0 %	0 %	0 %
Stadin Sote -uutiskirje on jälleen täynnä mielenkiintoisia sote-aiheita.	SERVICES	0 %	0 %	99 %	0 %	0 %	0 %
Helsingin kaupungin merellinen strategia ja toimintaohjelma on saanut kaupunginhallituksen hyväksynnän. Merellinei	SERVICES	0 %	0 %	100 %	0 %	0 %	0 %
Uusi Ilmastovisas alue -suunnitteluyökalu tukee liikenneasemien ilmastoystävällistä kaupunkisuunnittelua. Helsingii	SERVICES	2 %	10 %	84 %	4 %	0 %	0 %
Helsingin kaupunki on kehittänyt kansainvälisestikin ainutlaatuisen ilmastotyön seurantarapalvelun, Ilmastovahdin, pää	SUSTAINABILITY	0 %	99 %	1 %	0 %	0 %	0 %
Helsinki haluaa olla maailman toimivin ja parhaiten digitalisaatiota hyödyntävä kaupunki maailmassa. Miten tämä oike	FUNCTIONALITY	95 %	0 %	5 %	0 %	0 %	0 %
Kuvassa: Helsingin merellisen strategian projektipäällikkö Minttu Perttula avusti kaupunkisuunnittelun rantaautumis	SERVICES	0 %	0 %	100 %	0 %	0 %	0 %
Helsingin kaupunginosien stadiuutit ovat aloittaneet työnsä. Lue lisää ohesta!	FUNCTIONALITY	99 %	0 %	0 %	0 %	1 %	0 %
Oiva sää puistojumppalle! Tähän liikutaa mm. Latokartanossa, Rastilassa ja Töölönlahdella. Kaikki kaupungin mak	SERVICES	0 %	0 %	100 %	0 %	0 %	0 %
Haluamme tuoda palvelut lähemmäksi ihmisten arkea ja tehdä siitä entistäkin sujuvampaa. Auta meitä Helsingin kau	SERVICES	5 %	1 %	94 %	0 %	0 %	0 %
Tänä kesänä Helsingissä voi nauttia merestä kaupunkiveneellä soutelemalla, kun vuokravälitysyritys Skipperi aloitt	SERVICES	0 %	0 %	100 %	0 %	0 %	0 %
Messukeskuksen katolle on valmistunut aurinkovoimala. Aurinkovoimala käynnistetään elo-syyskuun vaihteessa. S	SUSTAINABILITY	0 %	100 %	0 %	0 %	0 %	0 %
Helsingin kaupunginhallitus hyväksyi kaupungille uudet eettiset periaatteet. Maailman toimivin kaupunki – Helsingin k	FUNCTIONALITY	99 %	0 %	0 %	0 %	0 %	0 %
Miten sinä käyttäisit 4,4 miljoonaa euroa? Tutustu osallistuvaan budjetointiin: https://omastadi.hel.fi/	MANAGEMENT	0 %	0 %	0 %	100 %	0 %	0 %
Reittioppaasta saat aikataulut ja lähtöajurit yhteensä 26 pääkaupunkiseudun saareen, jotka ovat säännöllisen vesii	SERVICES	0 %	0 %	100 %	0 %	0 %	0 %
Helsingin kaupunki haluaa toteuttaa merellistä strategiaa lisäämällä Kaunissaaren vetovoimaisuutta. Kaunissaareen	SERVICES	0 %	0 %	100 %	0 %	0 %	0 %
Helen parantaa kaukolämpöverkon energiatehokkuutta hankkimalla kaukolämpöverkkoa optimoivan digitaalisen sää	SERVICES	0 %	6 %	94 %	0 %	0 %	0 %
Suomen EU-puheenjohtajakauden teemana on: Kestävä Eurooppa – kestävä tulevaisuus. Pääkaupunkina Helsinki i	LOBBYING	0 %	1 %	0 %	0 %	98 %	0 %
Rakentamisen kiertotaloutta edistetään muun muassa kokeilin pääkaupunkiseudulla kansainvälisessä EU:n rahoitt	SUSTAINABILITY	0 %	99 %	0 %	1 %	0 %	0 %
Helsinki haluaa olla maailman toimivin kaupunki. Toimivuus tarkoittaa etenkin sujuvaa arkea. Digitalisaation avulla k	FUNCTIONALITY	97 %	0 %	2 %	0 %	0 %	0 %

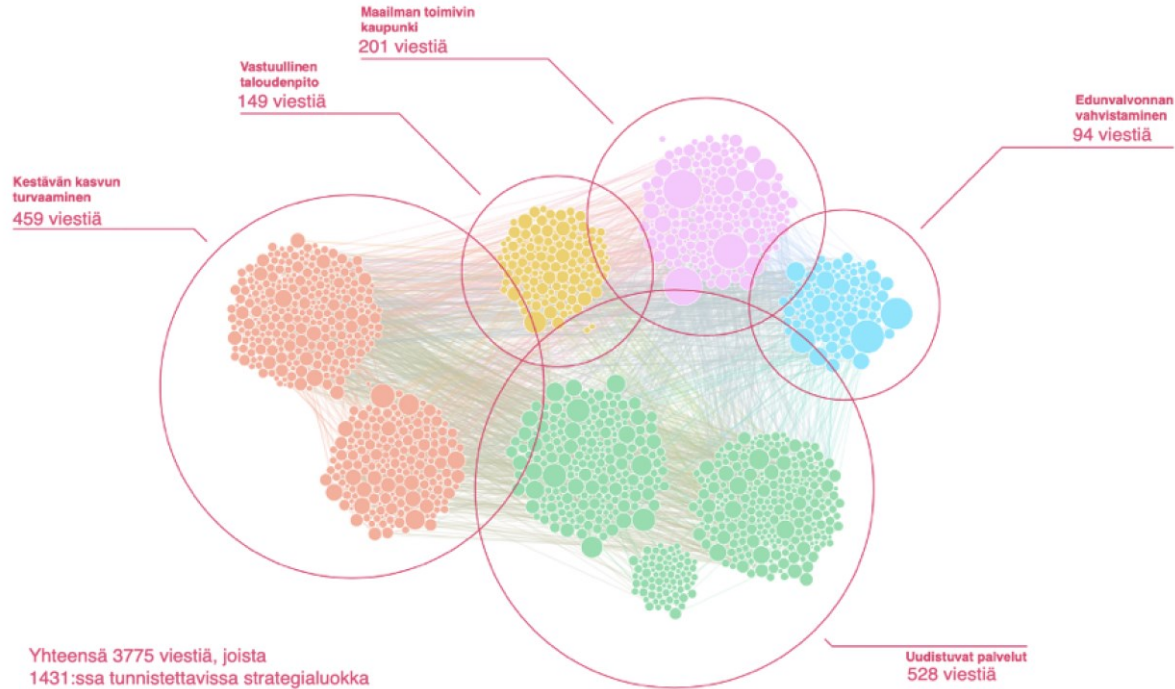
Helsinki

Oikein luokiteltu

Osittain oikein

Väärin luokiteltu

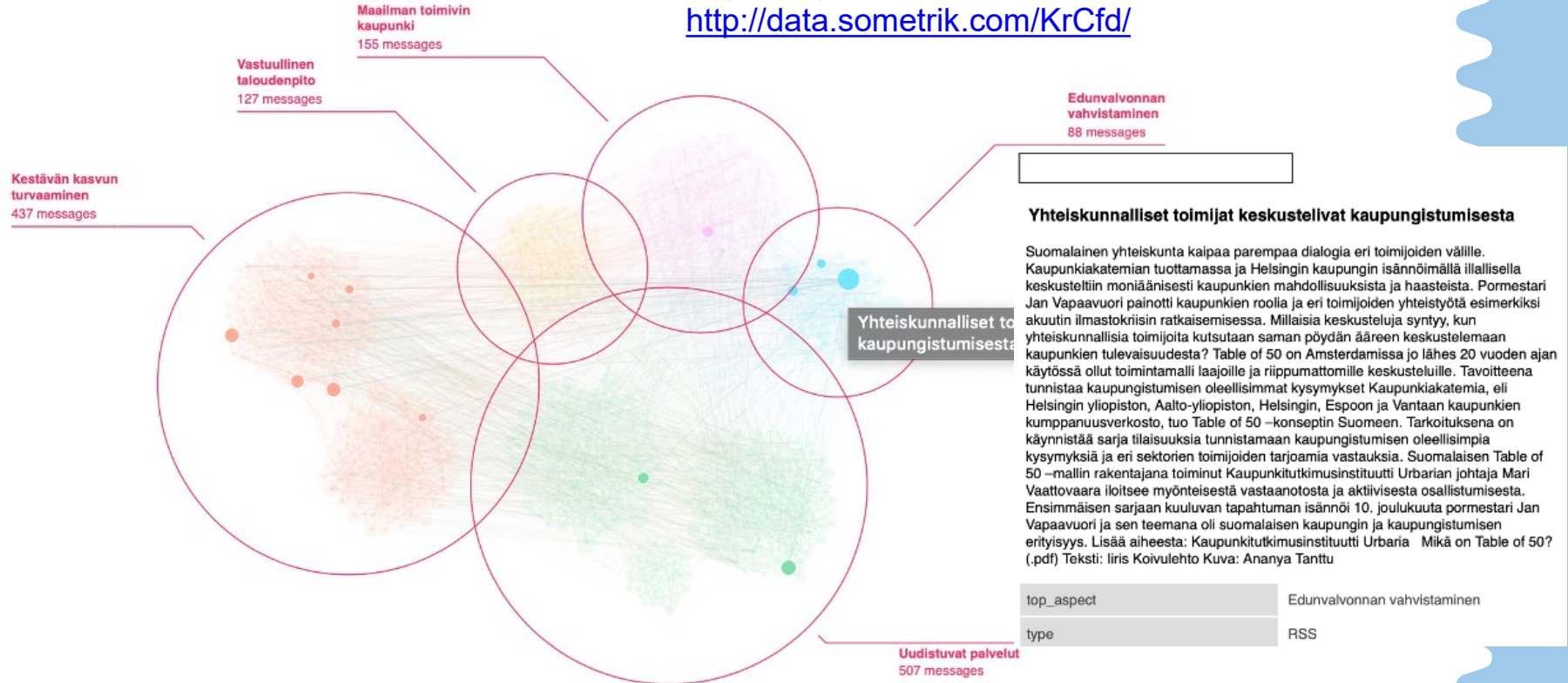
3. Kokeilun tuotokset: visualisointi



3. Kokeilun tuotokset: käyttöliittymä

Käyttöliittymän avulla mahdollista porautua aineistoon:

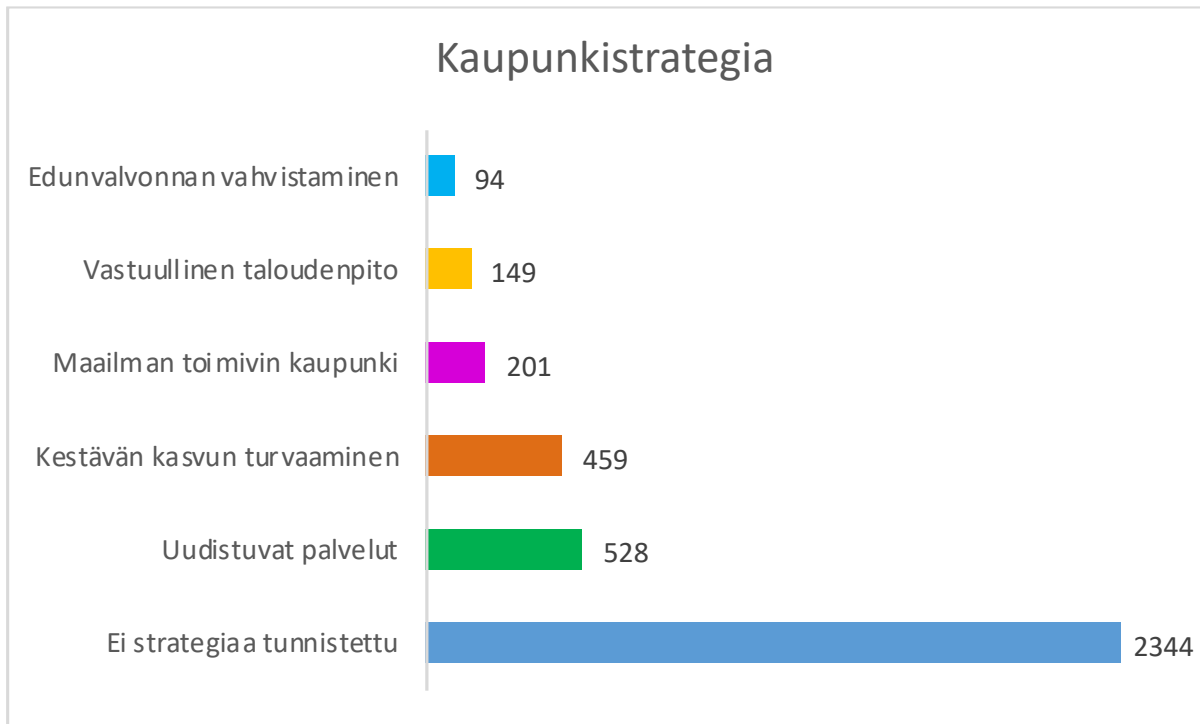
<http://data.sometrik.com/KrCfd/>



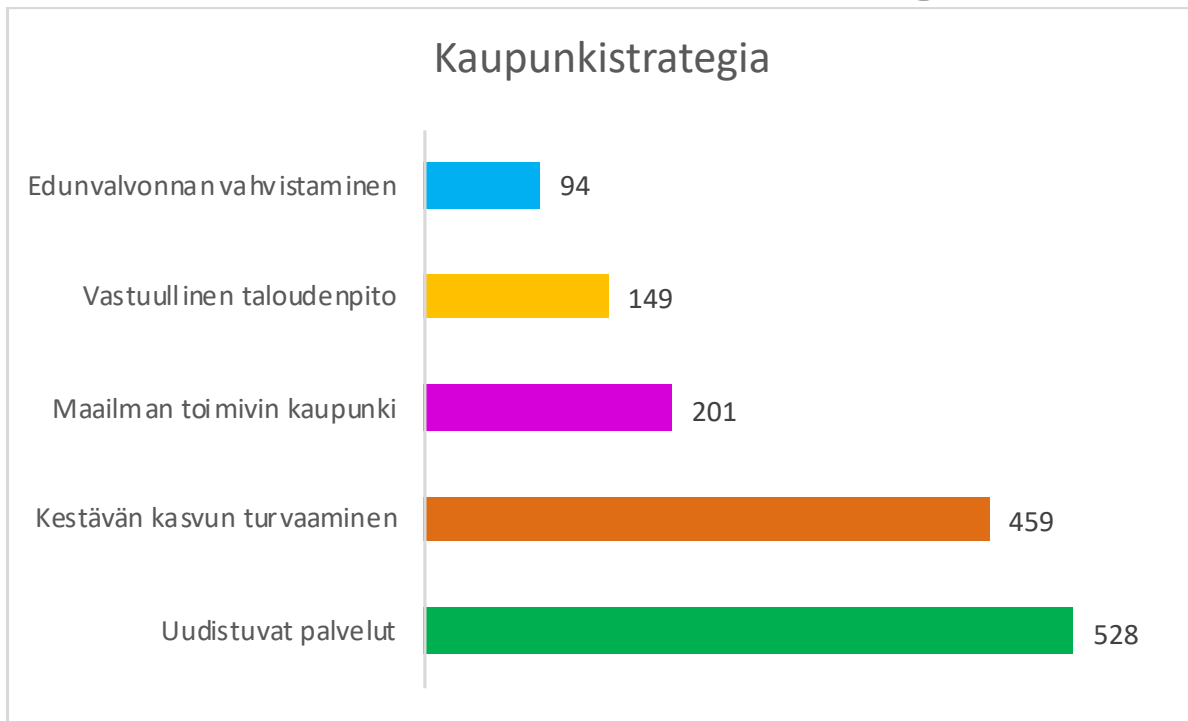
Esimerkki strategian valinnasta

Otsikko	Teksti	Strategia	Maailman toimivin kaupunki	Kestävän kasvun turvaaminen	Uudistuvat palvelut	Vastuullinen taloudenpito	Edunvalvonnan vahvistaminen	Ei strategiaa
Yhteiskunnalliset toimijat keskustelivat kaupungistumisesta	Suomalainen yhteiskunta kaipaa parempaa dialogia eri toimijoiden välille. Kaupunkiakatemian tuottamassa ja Helsingin kaupungin isännöimällä illallisella keskusteltiin moniäänisesti kaupunkien mahdollisuuksista ..	Edunvalvonnan vahvistaminen	3 %	15 %	1 %	1 %	44 %	36 %
Helsinki kasvattaa panostustaan kansainvälisten osaajien houkutteluun	Helsinki satsaa nyt kaupungin kansainvälisen maineen rakentamiseen yhteistyössä teknologia-alan yritysten kanssa. Syyskuussa käynnistyvän työn tavoitteena on testata, millaisella mielikuvalla kansainvälisiä osaajia saadaan pysyvästi houkutteltua muuttamaan Helsinkiin...	Edunvalvonnan vahvistaminen	21 %	16 %	0 %	1 %	61 %	1 %

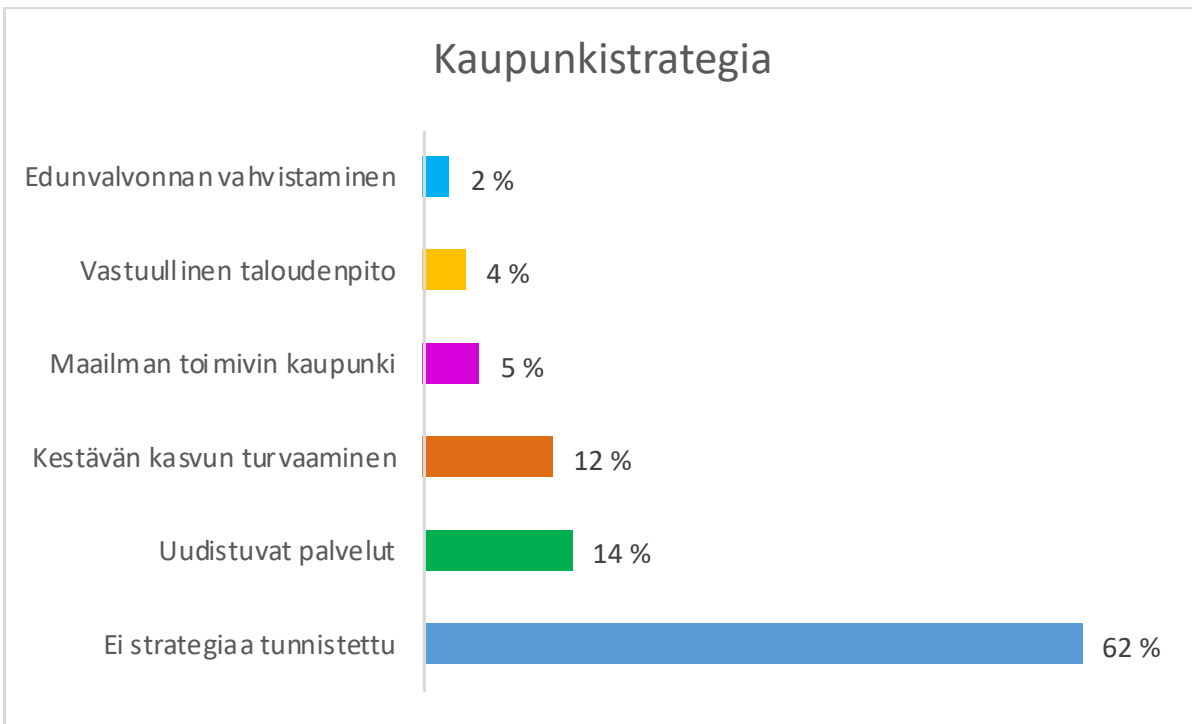
Kaupunkistrategia: luokkien määrät koko aineistossa



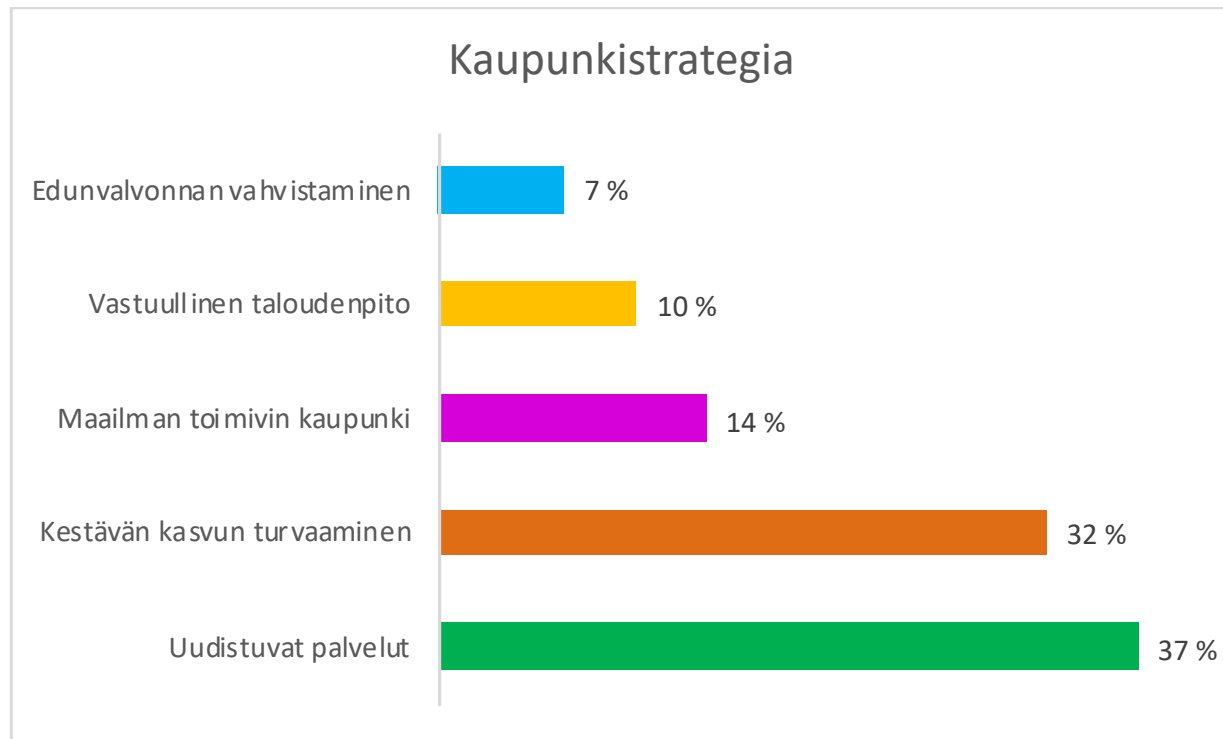
Kaupunkistrategia: viestit, joissa tunnistettavissa strategia



Kaupunkistrategia: luokkien osuus koko aineistosta



Kaupunkistrategia: Osuus aineistosta, jossa tunnistettavissa strategia



4. Opit tekoälyn mahdollisuuksista

Ihmisen tekemä analyysi

- Ihminen pystyy käsittelemään vain rajallisen määrän tietoa (tässä tekstiä).
- Analyysi perustuu usein otokseen tai mutuun

Tekoäly analyysin apuna

- Tekoälyllä voidaan hyödyntää laajempia tietomääriä
- Tekoälyn avulla voidaan analysoida dataa reaaliaikaisesti
- Tekoäly luokittelee johdonmukaisesti

Tekoälyn heikkoudet / ominaispiirteet

- Tekoäly kuvastaa sitä opettavan ihmisen näkemyksiä
- Vinoumat opetusdatassa näkyvät tekoälyn toiminnassa

Tekoälyn soveltaminen

- Tekstinluokittelu luo pohjan muulle soveltamiselle
- Tietojen yhdistäminen muun tiedon kanssa

5. Opit tekoälyn kehittämisestä

Osaaminen opetusdatan luonnissa

- Määrittely ja ohjeet ovat tarpeen opettamisessa
- Substanssi pitää olla hyvin hallussa. Tarvitaan perehdytystä / tuntemusta itse aiheesta
- Kaupunkistrategian määrittely on hyvin haastavaa.
 - Strategia-kulma vs. yleinen aiheeseen liittyvä asia. Esim. Hyvinvoinnin ja terveyden edistämisen johtamisrakenne vs. terveydenhoito yleisesti
 - Osa teemoista on yleisiä, osa taas yksityiskohtaisia.
 - Luokat eivät ole yksiselitteisiä
 - Luokittelu oli organisaatiolähtöistä
- Luokittelijan olisi hyvä tietää käyttötarkoitus ennalta. Tämä auttaisi luokkien hahmottamisessa.
- Luokittelijoiden yhteinen kickoff-opetussessio on tarpeen yhteisen kielen löytämiseksi
- Tekstien palasteleminen auttoi opetuksessa

5. Opit tekoälyn kehittämisestä

Opetusdata ja mallin koulutus

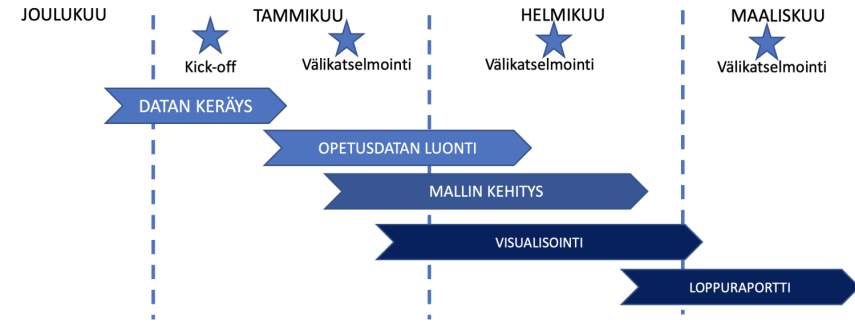
- Opetusdataa tarvitaan paljon per luokka
- Osasta luokista ei ollut aluksi tarpeeksi opetusdataa
 - Luokkien yhdistämistä myös kokeiltiin
 - Lopullisessa mallissa on käytössä päästrategiat, jotka ovat lähtökohtaisesti laajempia, kuin alkuperäiset luokat
- Mallia opetettiin rinnan opetusdatan luonnin kanssa
- Ensimmäinen vaihe: opetusdatana koko dokumentti
 - Ongelmana tekstin ”yleisyys”
- Seuraava vaihe: pilkottiin opetusdata kappaleisiin
 - Kohdennetumpaa opetusdataa

6. Opit kokeilemisesta

- Ketteryyttä ja kokeilumentaliteettia vaaditaan kaikilta toimijoilta
- Tavoitteen määrittely tärkeää.
Kokeilussa lähdettiin kokeilemaan ilman kirkasta maalia siitä, mikä on lopullinen käyttötarkoitus
- Aiheena kaupunkistrategia on laaja
- Mieluummin luokitellaan enemmän ja tarkemmin samalla vaivalla.
Ts. eri luokittelun ulottuvuudet kerralla.
- Kokeilu on iteratiivinen prosessi. Maali ja suunnitelmat voivat muuttua matkan varrella
- Yhteistyö luokittelijoiden välillä on tärkeää yhteisen kielen löytymiseksi sekä parhaiden käytäntöjen jakamiseksi
- Feedback-loop-tärkeää. Koneen luokittelemaa tietoa pitää voida
arvioida.

7. Opit resursoinnista

- Kokeilu pysyi aikataulussaan
- Kokeilu pysyi budjetissaan
- Koneen opettamiseen varattava myös aikaa
 - Luokittelu vaatii paljon aikaa!
- Parhaita käytäntöjä resursointiin
 - Hajauttaminen olisi hyvä
 - Pitää olla asiantunteva resurssi
 - Ihmisten motivointi
 - Luokittelun helpottaminen teknisesti. Esim. kone ehdottaa luokkaa, jonka luokittelija hyväksyy tai hylkää
 - Monta luokittelijaa mahdollistaa paremman laadun ja pidemmän vinouman



8. Kokeilun tekninen ympäristö

Datan hankinta ja esikäsittely

- Ratkaisussa datan hankinta tapahtui Helsingin kaupungin RSS-syötteestä kerätyllä aineistolla, Helsinki-lehden html-versiosta sekä Helsingin kaupungin Facebook-sivujen ylläpitäjän omista julkaisuista
- Aineiston dokumentit pilkottiin kappaleiksi, ja kappaleet syötettiin Sometrikin omaan tarkoitukseen kehitettyyn, selaimessa toimivaan luokitteluopetusympäristöön, jossa luokittelua tekevät ihmiset saivat antaa mahdolliset luokat, joihin kappale kuuluu. Tätä ennen kokeiltiin koko dokumentin mittaista luokittelua, mutta tämän luokittelijat kokivat hankalaksi ja hitaaksi.
- Mallin rakentamista varten kaikki data esikäsiteltiin, esim. sanat muutettiin perusmuotoon (Helsingin -> Helsinki) ja hälysanat (ja, ei, on jne.) poistettiin dokumenteista.



8. Kokeilun tekninen ympäristö

Koneoppimismalli

- Tekoälyksi projektia varten valittiin FastText.
FastText valittiin sen nopeuden ja hyvien kokemusten perusteella. FastText on Facebookin vuonna 2015 julkaisema ja jatkuvasti kehittämä kirjasto. FastText rakentaa sana-avaruuden mallin neuroverkkojen avulla. Kielinä C++.
- Malli rakennettiin Sometrikin omalla serverillä.
Kaikki data syötettiin FastText-kirjastoon koko sana-avaruuden (word embedding) kielimallin rakentamiseksi. Tähän malliin syötettiin myös pienempi, luokiteltu aineisto, josta luokittelija rakennettiin.

SOMETRIK

*fast*Text



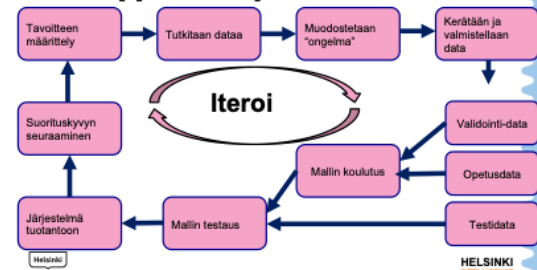
Facebook
Open Source

8. Kokeilun tekninen ympäristö

Mallin rakennus ja visualisointi

- Koko aineisto syötettiin luokittelijaan, jolloin saatiin ensimmäinen aineisto arvioitavaksi.
- Aineistosta poimittiin väärin luokiteltu aineisto ja mallin rakennus tehtiin uudelleen.
 - Nämä iteraatiot toistettiin kahteen kertaan.
- Koneen arvioima aineisto muutettiin verkostomuotoon, josta visualisointi rakennettiin.
- Visualisoinnista tehtiin D3.js-muotoinen versio selainpohjaiseen käyttöliittymää varten.

Koneoppimisen prosessi



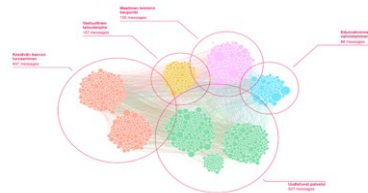
8. Kokeilun tekninen ympäristö

Tuotteistus ja kehitysmahdollisuudet

- Lopullinen malli toimii Sometrikin serverillä, johon voidaan ajaa uutta aineistoa sisään esim. API:n kautta.
- Käyttöliittymä: <http://data.sometrik.com/KrCfd/>

Mitä teknisessä ympäristössä olisi hyvä ottaa huomioon, jos kokeilu tehtäisiin nyt uudelleen?

- Opetusympäristöön toivottiin mahdollisuutta maalata tekstistä luokkaa vastaava kohta



Paragraph in Helsingin kaupunki

Me Työterveydessä olemme olemassa, jotta Helsinki voi hyvin. Tähän sanomaan voi hyvin samaistua, ja tuntea ylpeyttä omasta työstään ja kaikesta siitä, mitä työme tekee mahdolliseksi. Vain hyvinvoivat työntekijät voivat tehdä Helsingistä maailman toimivimman kaupungin.

Helsingin kaupunki | unknown at 8.11.2019 10:16:13 | 87732 fi

- ☐ Toimiva Maailman toimivim kaupunki
- ☐ Kestävä kasvu Kestävän kasvun turvaaminen
- ☐ Palvelut Uudistuvat palvelut
- ☐ Taloudenpito Vastuullinen taloudenpito
- ☐ Edunvalvonta & yhteistyö Edunvalvonta | Yhteistyö
- ☐ Ei strategiaa Ei strategiaa

9. Kokeilun data

- Data oli julkisesti saatavilla olevaa tekstimuotoista dataa
 - Hel.fin uutiset:
 - <https://www.hel.fi/uutiset/fi>
 - Facebook-tilin postaukset:
 - @helsinginkaupunki
 - Helsinki-lehti
 - <https://www.hel.fi/kanslia/helsinkilehti-fi/uutiset/>
 - Helsinki-lehden PDF-versiot jätettiin pois kokeilusta
 - <https://www.hel.fi/kanslia/helsinkilehti-fi/arkisto/>
- Datan saatavuuteen on hyvä panostaa ennalta, jotta data helposti käytettävää eri käyttötarkoituksiin
- Datan omistajuus pitää olla aina kaupungilla (tai julkista)

10. Jatkopäätökset ja -ideat

- Vastaavasti voitaisiin toteuttaa asiasanoitus sisällöntuotannon prosessiin
 - Teema, paikka, henkilö (NER-analyysi).
- Sopisi myös työkaluksi viestinnän seurantaan
 - Kokeilu lisäsi kiinnostusta työkaluun jolla tekstiä pystyy louhimaan
 - Sosiaalisen median ja uutismedian seuranta auttaisi asiakaslähtöiseen sisällöntuotantoon
- Uusia datalähteitä hyödynnettäväksi?