

Opinnäytetyö AMK

Liiketalous

Markkinointi urheilijoille

2015

Hannes Kari

# MARKKINATUTKIMUS AVOIMEN TEKSTIN ANALYSOINTITYÖKALULLE



TURUN AMMATTIKORKEAKOULU  
TURKU UNIVERSITY OF APPLIED SCIENCES

OPINNÄYTETYÖ (AMK) | TIIVISTELMÄ

TURUN AMMATTIKORKEAKOULU

Liiketalous | Markkinointi urheilijoille

Valmistumisvuosi 2015 | Sivumäärä: 30

Olli-Pekka Lehtisalo

Hannes Kari

## MARKKINATUTKIMUS AVOIMEN TEKSTIN ANALYSOINTITYÖKALULLE

Tämän opinnäytetyön taustalla oli Analystica Oy:n halu kehittää markkinoille kustannustehokas tapa analysoida suuria määriä tekstejä. Analystican avoimen tekstin työkalu oli jo opinnäytetyötä aloitettaessa kehitysvaiheessa. Tämän työn tavoitteena oli selvittää työkalun tarvetta ja mahdollisuuksia markkinoilla sekä tutkia yleisimpiä ongelmia, joita vastaavanlaisilla työkaluilla esiintyy. Tavoitteena oli kerätä tietoa kolmelta eri kohderyhmältä, joilla arveltiin olevan tarvetta kyseiselle työkalulle.

Markkinatutkimus suoritettiin niin, että lähetettiin sähköpostilla kutsu osallistua haastatteluun tarkkaan valituille yritysten edustajille. Päätoiveena oli saada suorittaa puhelinhaastattelu, mutta mahdollisuudeksi annettiin myös vastata nettikyselyyn sekä mahdollisuus siirtää kutsu eteenpäin muulle yrityksen tai organisaation edustajalle. Näin toimittiin, jotta saataisiin mahdollisimman suuri määrä vastauksia ja luotettavat tutkimustulokset. Sekä haastattelun rakenne että nettikyselyn rakenne vastasivat samaa kaavaa.

Tutkimuksen tuloksena voidaan todeta, että Analystica on oikeilla jäljillä työkalunsa kehityksen suhteen. Kehitteillä olevat ominaisuudet herättivät kiinnostusta vastaajissa sekä tarvetta tällaiselle työkalulle on selkeästi olemassa, kunhan sen käyttö saadaan kustannustehokkaaksi. Opinnäytetyötä voidaan pitää onnistuneena, sillä tutkimuksen päätavoitteena olleen, työkalun markkinamahdollisuuksien tutkimisen lisäksi saatiin selville vinkkejä siitä, miten työkalua kannattaa markkinoida, eli mikä ihmisiä kiinnostaa sekä saatiin kerättyä useita mahdollisia ehdokkaita työkalun maksuttomaan kokeiluun varten.

### ASIASANAT:

Tekstin analysointi, laadullinen tutkimus, big data, tekstinlouhinta, markkinatutkimus, sisällönanalyysi

BACHELOR'S THESIS | ABSTRACT

TURKU UNIVERSITY OF APPLIED SCIENCES

Business administration | Marketing for athletes

Completion year of the thesis: 2015 | Total number of pages: 30

Olli-Pekka Lehtisalo

Hannes Kari

# MARKET RESEARCH FOR OPEN TEXT ANALYSIS TOOL

This thesis work started with Analystica Ltd.'s desire to develop for the market a cost efficient tool for analyzing large amounts of text. The development process for the open text analysis tool had already begun when this thesis work was started. The goal of this thesis was to find out the needs and possibilities of this tool in the market and to investigate the most common problems with the same kind of tools. The goal was to gather information from three different target groups whom all was believed to have the need for this kind of tool.

The market research was conducted by sending an invite for specific group of workers of different companies to participate in an interview. The main goal was to conduct a phone interview but the opportunity was given to either answer a web form or shift the invite to another person in the company. These measures were taken to ensure that there would be as many responses as possible and that the results would be reliable enough. Both the structure of the interview and the structure of the web form were designed in the same way.

As a result for this research it can be noticed that Analystica is in the right track with the development of its tool. The features being developed gathered interest among the respondents and it was clear that there is a need for this kind of tool as long as it will be cost-efficient. This thesis work can be graded as successful because in addition to reaching the main goal, there were indications how the tool could be marketed and Analystica gathered several possible candidates for the free experiment of the tool.

## KEYWORDS:

Text analysis, qualitative research, big data, text mining, market research, content analysis

# SISÄLTÖ

<b>KÄYTETYT LYHENTEET (TAI SANASTO)</b>	<b>6</b>
<b>1 JOHDANTO</b>	<b>7</b>
<b>2 MARKKINATUTKIMUS</b>	<b>8</b>
2.1 Laadullinen tutkimus	9
2.2 Laadullisten tutkimusten käyttö nykypäivänä	10
2.3 Sisällönanalyysi	11
<b>3 BIG DATA</b>	<b>12</b>
3.1 Tavaroiden internet	13
3.2 Strukturoitu ja stukturoimaton data	13
3.3 6Aika	14
3.4 Avoin data	14
<b>4 AVOIMEN TEKSTIN ANALYSOINNIN TYÖKALU</b>	<b>15</b>
4.1 Ongelmien ratkaiseminen	17
4.2 Vertailumahdollisuudet	18
4.3 Tiedon visualisointi	19
<b>5 TUTKIMUS</b>	<b>21</b>
5.1 Menetelmä	21
5.2 Sähköpostikutsut	22
5.3 Kohderyhmä	23
5.4 Puhelinhaastattelut	23
5.5 Internetkyselylomake	24
<b>6 TULOKSET</b>	<b>26</b>
6.1 Tutkimuksen luotettavuus	27
6.2 Pohdinta	28
<b>LÄHTEET</b>	<b>30</b>

## KUVAT

Kuva 1. Esimerkki sanapilvestä.	13
Kuva 2. Eri puolueiden vaalilupauksien vertailu.	16

## KÄYTETYT LYHENTEET (TAI SANASTO)

Asiasanapilvet	Asiasanapilvet, avainsanapilvet tai sanapilvet ovat tekstistä poimittujen yleisimpien sanojen kokoelma. Mitä suurempana sana pilvessä näkyy, sitä useammin se tekstissä esiintyy.
Avoin data	Avoimella datalla tarkoitetaan, kenelle tahansa maksutta ja luvallisesti uudelleen käytettävissä olevaa dataa, joka on kone-luettavassa muodossa.
Data	Tieto, joka on koneellisesti käsiteltävässä muodossa.
Lemmatsointi	Eli normalisointi tai perusmuotoistaminen tarkoittaa aineiston muuttamista perusmuotoon. Esimerkiksi kissa, kissan ja kis-salle kaikki muuttuvat sanaksi kissa.
Tiedonlouhinta	Eli data mining tarkoittaa joukkoa menetelmiä, joilla pyritään oleellisen tiedon löytämiseen suurista tietomassoista.

# 1 JOHDANTO

Kvalitatiivisen eli laadullisen tutkimuksen tekeminen on yleensä haastavaa ja aikaa vievää. Toisin kuin kvantitatiiviseen eli määrälliseen tutkimukseen, laadulliseen tutkimukseen on käytössä vain rajoitettu määrä työkaluja ja niissä vain rajoitettu määrä toimintoja, jotka auttavat tulosten analysointia.

Määrällisen tutkimuksen tekeminen helpottuu usein huomattavasti, mikäli vertailupohjana on yksi tai useampi laadullinen tutkimus, jotka selvittävät tutkittavan asian trendejä ja suuntauksia. Näiden avulla määrällisen tutkimuksen kysymyspatteriston rakentaminen helpottuu huomattavasti.

Keväällä 2015 aloitin työharjoittelun turkulaisessa tietualan pienyrityksessä Analystica Oy:ssä. Pääsin mukaan projektiin, jonka tarkoituksena oli rakentaa työkalu avoimen tekstin laadukkaaseen analysointiin. Työkalun tarkoituksena on helpottaa suurten tekstimassojen analysointia, mikä yleensä on ongelmallista ja aikaa vievää esimerkiksi laadullisessa tutkimuksessa.

Projekti oli vasta alkuvaiheessa, joten tehtäväkseni tuli tehdä markkinatutkimusta mahdollisille asiakasyrityksille. Tarkoituksena oli saada selville asiakasyritysten tarpeita sekä selvittää yleisimpiä ongelmia, joita jo käytössä olevissa työkaluissa on. Näiden pohjalta on tarkoitus luoda työkalu, jollaista markkinoilla ei vielä ole.

Tässä opinnäytetyössä selvitetään teorian pohjalta kyseisen työkalun mahdollisuuksia sekä kerrotaan markkinatutkimuksella saavutetuista tuloksista. Lisäksi tässä työssä käsitellään, miten projekti eteni tutkimusten johdattelemana.

## 2 MARKKINATUTKIMUS

Markkinatutkimuksella, kuten nimestäkin voi päätellä, tutkitaan tietyn alan tai tuotteen markkinoita. Tarkoituksena on saada laajempi ymmärrys, miten tutkittu markkina-alue käyttäytyy. Jotta tähän kyettäisiin, tulee kerätä mahdollisimman monipuolinen datamassa kyseisestä markkina-alueesta.

Markkinatutkimuksessa yhtä tärkeää on myös ymmärtää, miksi se tehdään eli mitä tutkimuksella tavoitellaan. Tavoitteita voi olla erilaisia, ja markkinatutkimus tuleekin suunnitella tavoitteen mukaisesti.

Tutkimustulosten analysointi on myös tehtävä huolella, jotta kerätystä aineistosta saadaan maksimaalinen hyöty. Epämääräisesti suunnitellusta, suoritetusta tai analysoidusta markkinatutkimuksesta saattaa olla enemmän haittaa kuin hyötyä.

Markkinatutkimuksen syvin tarkoitus on auttaa päätöksenteossa. Hyvin suoritetun tutkimuksen avulla pystytään perustelemaan tarkasti tehdyt päätökset sekä välttämään riskejä. Tämän vuoksi onkin ymmärrettävä jo tutkimusta suunniteltaessa, millaisiin kysymyksiin haetaan vastausta.

Markkinointitutkimus ja markkinatutkimus eivät tarkoita samaa asiaa, mutta markkinatutkimusta voidaan hyödyntää myös markkinointipäätöksissä. Tuotteen tai tuoteryhmän markkinoinnissa on tärkeää ymmärtää oman tuotteen heikkoudet ja vahvuudet. Näiden ymmärtämisen apuna käytetään usein markkinatutkimusta.

Markkinatutkimuksella voidaan myös esimerkiksi selvittää soveltuuko jokin tuote lainkaan markkinoille, löytyykö sille tarvetta ja kiinnostusta. Joidenkin prototyyppien massavalmistus saatetaan unohtaa markkinatutkimuksella kerättyjen tietojen perusteella. Tällöin voidaan välttää isojakin katastrofeja.

Markkinatutkimuksia voidaan yhtä lailla hyödyntää myös jo olemassa olevien tuotteiden markkinoiden tutkimiseen ja sitä kautta tuotteiden kehittämiseen markkinoille sopivammiksi. Useat yritykset suorittavatkin markkinatutkimuksia samoista aiheista uudelleen kerätäkseen ajankohtaista tietoa markkinatilanteesta. Mikäli tutkimukset suoritetaan jotakuinkin samalla tavalla, ovat ne myös ainakin



osittain verrattavissa toisiinsa. Tällöin on mahdollista havaita tiettyjä trendejä sekä kehittää ennusteita, mihin markkinatilanne on tulevaisuudessa kehittymässä. (Hague & Jackson 1999, 11–17)

Erilaisia markkinatutkimuksen tiedonkeruukanavia ovat mm.

- Kirjepostitus, jossa tutkimuslomake
- Sähköpostikysely
- Web-lomake
- SMS-kysely
- Kasvotusten tapahtuva haastattelu
- Puhelinhaastattelu
- Tablet-tutkimus

Tiedonkeruutapa tai -tavat tulee valita valitun kohderyhmän mukaisesti. Ei välttämättä esimerkiksi ole mielekästä suorittaa usean vastaajan tutkimusta kasvotusten haastattelun. Tiedonkeruutavan valintaan vaikuttaa myös suoritetaanko tutkimus laadullisena vai määrällisenä tutkimuksena.

## 2.1 Laadullinen tutkimus

Laadullisen tutkimuksen käsitteen määrittely on hankala. Pohjimmiltaan kyse on siitä, mistä näkökulmasta asiaa katsotaan. Laadullinen tutkimus voi edustaa joko teoreettista tai empiiristä näkökulmaa. Laadullinen tutkimus vaatii joka tapauksessa teoriapohjaa, mutta näkökulmasta riippuen tutkimus voidaan toteuttaa teoreettisena tai empiirisenä.

Tyypijako teoreettiseen ja empiiriseen on kuitenkin ongelmallinen, sillä osa laadullisista tutkimuksista voidaan sijoittaa molempiin kategorioihin. Yhteenvetona voidaan kuitenkin todeta, että laadullinen tutkimus ei voi omaksua teoriaa yksin itselleen, mutta se ei voi sitä myöskään kokonaan hylätä.

Empiirisessä näkökulmassa korostuvat aineiston keräämis- ja analyysimetodit. Näiden avulla lukijalla on mahdollisuus arvioida tutkimusta ja sen luotettavuutta. Tutkimusraportissa metodit on argumentoitava.

Teoreettisesta näkökulmassa sitä vastoin puuttuvat tarkat menetöt. Teoreettisessa analyysissä on kyse eräänlaisesta ongelmanratkaisuesseestä. Tällöin ratkaisevaksi muodostuu se, miten argumentointi suoritetaan.

Teoreettisessa analyysissä korostuu aina, kuka on sanonut mitä ja milloin. Empiirisessä analyysissä taas yksilöiden tunnistettavuudella ei ole merkitystä. Oikeastaan päinvastoin, yksilöllinen tunnistettavuus häivytetään tarkoituksella.

Useimmiten laadullinen tutkimus muistuttaa lähemmin empiiristä tutkimusta, vaikka tutkimuksen raportointi sisältäisikin joitain suoria lainauksia. Tämä saateen usein tehdä ajatellen sen lisäävän tutkimuksen luotettavuutta. (Tuomi & Sarajärvi 2008, 17–22)

## 2.2 Laadullisten tutkimusten käyttö nykypäivänä

Laadullinen eli kvalitatiivinen tutkimus on vahvistanut vuosi vuodelta asemaansa osana nykyaikaista markkinatutkimusta. Laadullisessa tutkimuksessa pääpainona on vastata kysymyksiin mitä, miksi ja kuinka. Laadullinen tutkimus toteutetaan usein kasvotusten ryhmäkeskusteluna tai henkilökohtaisina haastatteluina. Henkilökohtaiset haastattelut ovat myös mahdollista hoitaa puhelimitse.

Laadullista tutkimusta voidaan käyttää joko itsenäisesti tai kvantitatiivisen eli määrällisen tutkimuksen pohjana. Laadullisella tutkimuksella voidaan selvittää määrällistä tutkimusta varten tutkimusongelma tai muu aihealue, jota halutaan tutkia. Tämän avulla voidaan määrittää määrälliseen tutkimukseen oikeat kysymykset, jotta saadaan selville haluttuja kokonaisuuksia.

On myös mahdollista syventää määrällisestä tutkimuksesta saatuja tietoja laadullisella tutkimuksella. Tällöin määrällinen tutkimus on pohjana sille, mitä laadullisessa tutkimuksessa halutaan tutkia. (Taloustutkimus.fi 2015)

### 2.3 Sisällönanalyysi

Laadullisen tutkimuksen raskain ja aikaa eniten vievä vaihe on sisällönanalyysi. Sisällönanalyysi on perusmenetelmä, jota voidaan käyttää kaikissa laadullisissa tutkimuksissa. Sisällönanalyysi sisältää kolme osaa: luokittelu, tyypittely ja teemoittelu. Näiden kolmen osan avulla pystytään strukturoimaan aineisto, jotta on helpompi tehdä päätelmiä ja kirjoittaa yhteenveto. (Tuomi & Sarajarvi 2008, 91–93)

Sisällönanalyysin perimmäinen tarkoitus on saada yleiskuva arvioitavasta aineistosta. Tämä ainoastaan helpottaa johtopäätösten tekemistä, eli pelkkä sisällönanalyysi ei riitä kertomaan tuloksia laadullisesta tutkimuksesta. Sisällönanalyysi on tekstianalyysia, jossa etsitään tekstin merkityksiä. (Tuomi & Sarajarvi 2008, 103–104)

### 3 BIG DATA

Big data on tällä hetkellä suuresti pinnalla oleva käsite. Sillä tarkoitetaan tiedon määrän jatkuvaa eksponentiaalista kasvua. Siihen liittyy varsin vahvasti erilaisten tapojen etsiminen, joilla voisi hyödyntää tätä tietoa. Big data –käsitteellä ei ole vielä mitään vakiintunutta suomennosta. Yleisin lienee tällä hetkellä suora suomennos, iso data.

Käsite on erittäin moniselitteinen. Yhtä mieltä voidaan kuitenkin olla siitä, että datan määrä kasvaa jatkuvasti kiihtyvällä voimalla. Datan määrää tai sen kasvunopeutta on kuitenkin mahdoton mitata kunnolla, sillä molemmat muuttuvat jatkuvasti. (Salo 2013, 10)

Big datan on arvioitu käsitteenä syntyneen jo 2000-luvun alussa. Tunnetuksi se tuli vuoden 2005 tietämillä, mutta vasta 2011 se tuli todellisesti pinnalle (Salo 2014, 26). Suomessa käsite on edelleen murrosvaiheessa. Vaikka se onkin tunnettu käsite alalla työskenteleville ihmiselle, on se suurelle yleisölle vielä hieman tuntemattomampi. Kesällä 2013 tehdyssä tutkimuksessa Ivorio sai selville, että julkishallinnon tavoitteena on suurimmaksi osaksi aloittaa aiheeseen liittyviä käytännön hankkeita vasta 2016. (Salo 2014, 34)

Big datan suurimmat haasteet liittyvät datan määrän kasvuun ja sisällön monipuolistumiseen. Tämä on vaikeuttanut suuresti oleellisen datan tunnistamista yhä kasvavan datamäärän joukosta. Lisäksi datan tarpeeksi nopea jalostaminen ja siihen reagointi synnyttävät haasteita. (Salo 2014, 6)

Yksi tärkeä vaikuttaja ilmiön pinnalle nousuun on tekniikan jatkuva kehitys. Tämä on mahdollistanut uusia ja parempia tapoja tallentaa jatkuvasti suurempia määriä mitä erilaisempaa dataa. (Salo 2013, 10–11)

### 3.1 Tavaroiden internet

Toinen tärkeä asia ilmiön pinnalle nousuun on tavaroiden internet. Käsite on suora suomennos englanninkielisestä käsitteestä Internet of Things (IoT). Tavaroiden internetillä tarkoitetaan tietoverkkoon kytkeytyneiden laitteiden määrän jatkuvaa kasvua. Verkottuneiden laitteiden määrä on tällä hetkellä arviolta 10 ja 20 miljardin välillä. Laitteiden määrä on kasvanut nopeassa tahdissa viime aikoina ja kasvun oletetaan jatkuvan myös tulevaisuudessa. (Salo 2013, 12)

### 3.2 Strukturoitu ja stukturoimaton data

Datan voi määritellä kahteen lohkoon. Strukturoitu data tarkoittaa esimerkiksi asiakastietoa, jossa kerrotaan asiakkaan yhteystietoja ja ostohistoriaa. Strukturoimaton data taas tarkoittaa esim. kaupan valvontakamerasta tallennettua videokuvaa.

Näiden kahden määritelmän raja on kuitenkin häilyvä. Tämän vuoksi jotkut käyttävät semistrukturoidun datan määritelmää sellaisesta datasta, jossa on piirteitä molemmista.

Datan strukturoinnin taso vaikuttaa paljolti siihen, miten dataa voidaan käsitellä ja miten helppoa se on. Syntyvän datan määrän jatkuva kasvu taas vaikeuttaa kaiken datan tallentamista. Tällä hetkellä suuntaus näyttää, että yhä suurempi osa syntyvästä datasta jää tallentamatta. Tämä johtaa myös siihen, että datan nopeasta käsittelystä on iso hyöty, koska se mahdollistaa nopean reagoinnin.

Mikäli datavirta pystytään käsittelemään heti, voidaan välttää raakadatan tallentaminen kokonaan ja hyödyntää ainoastaan käsiteltyä dataa. Datamäärien jatkuva kasvu luonee tilanteen, jossa kyky esikäsitellä datavirrat on merkittävä osa big datan hyödyntämisprosessia. (Salo 2014, 27)

### 3.3 6Aika

6Aika on Suomen kuuden suurimman kunnan yhteinen projekti. Kuutoskaupunkeja ovat Helsinki, Espoo, Vantaa, Tampere, Turku ja Oulu. Tavoitteena on parantaa näiden kaupunkien palveluita ja niiden toimivuutta yli kuntarajojen. Tämä on kansallisesti merkittävä hanke, sillä näiden kaupunkien alueilla asuu 30 % Suomen väestöstä.

Projektin kolmeen painopisteeseen kuuluvat avoin data, avoimet innovaatio-alueet sekä avoin osallisuus ja asiakkuus. Tarkoituksena on helpottaa kolmansien osapuolten osallistumista kaupunkien palvelujen kehittämiseen. Yhteistyön helpottamiseksi kaupungit pyrkivät avaamaan prosessejaan ja dataansa sekä kehittämään helpottavia työkaluja ja toimintamalleja. (<http://6aika.fi/kaupungit/>, 2015)

### 3.4 Avoin data

Yksi kolmesta 6Aika-projektin painopistealueista oleva avoin data on merkittävä projektin kannalta, sillä kaupunkien halukkuus avata dataansa yleiseen käyttöön, luo kasvavassa määrin uusia mahdollisuuksia datan hyötykäyttöön. Analystica Oy onkin hyödyntämässä 6Aika-projektia yhtenä uuden tekstin analysointi työkalunsa pilottina.

Datan avaamisen on tarkoitus luoda innovaatioita. Yritykset voivat käyttää dataa uusien palveluidensa raaka-aineena. Datan avaamisessa keskitytäänkin erityisesti yritystoimintaa hyödyntävän datan avaamiseen. (<http://6aika.fi/painopisteet/avoin-data/>, 2015)

## 4 AVOIMEN TEKSTIN ANALYSOINNIN TYÖKALU

Keväällä 2015 alkanut projekti pohjautuu Analystica Oy:n haluun kehittää työkalu suurien tekstimäärien helpompaan analysointiin. Kuten jo kappaleessa 2.1 mainittiin, pelkästään laadullisten markkinointitutkimusten määrä on ollut jatkuvassa kasvussa.

Ilmiö big data taas luo ainakin teoriassa uusia mahdollisuuksia ja tarpeita erilaisille datan analysointityökaluille. 6Aika-projektin yksi painopisteistä, datan avaaminen, helpottanee erilaisten aineistojen käsittelymahdollisuuksia.

Analystican ensimmäinen avoimen tekstin analysoinnin pilottiprojekti aloitettiin yhteistyössä yhden yrityksen suurimman asiakkaan, Helsingin ja Uudenmaan sairaanhoitopiirin (HUS), kanssa. Pilottiprojekti aloitettiin huhtikuussa 2015 HUS:n aineistoja tutkimalla.

Yhdeksi työkalun tärkeimmäksi analysointitavaksi on tarkoitus kehittää asiasanapilvet (kuva 1.). Asiasanapilvien, avainsanapilvien tai ainoastaan sanapilvien tarkoituksena on luoda käyttäjälle käsitys, mistä asioista tutkittavassa aineistossa puhutaan.



Kuva 1. Esimerkki sanapilvestä.

Sanapilvien toiminta perustuu lyhyesti siihen, että käytössä oleva ohjelma louhii haluttua aineistoa tuoden esille aineistossa yleisemmin esiintyvät sanat. Alkeellisia avainsanapilviä pystytään jo nyt kehittämään jopa ilmaiseksi internetissä olevien palveluiden avulla. Niiden hyödyllisyyttä voi kuitenkin pitää erittäin kyseenalaisena.

Katleena Korteso (2013) toteaa blogissaan Ei oo totta avainsanapilvissä olevan viisi merkittävää ongelmaa:

- 1) Aineistoista ei useinakaan ole valmiiksi poistettu yleisempiä suomen kielen sanoja. Tällöin sanapilvessä on sanoja, kuten *ja*, *ei*, *se*, jotka eivät anna minkäänlaista lisäarvoa tutkittavasta aineistosta.
- 2) Sanapilvet eivät tunnista taivutusmuotoja. Suomen kielen monet taivutusmuodot tekevät sanapilvistä sekavia, ja epäkäytännöllisiä
- 3) Avainsanapilvet eivät tunnista synonyymejä ja sanojen eri variantteja. Tällöin esim. *surkea*, *huono*, *surkeahko* ja *kurja* lasketaan eri sanoiksi, vaikka useissa tilanteissa tarkoitusperä on sama.
- 4) Avainsanapilvet eivät ota huomioon sanaparien ja sanaryhmien merkityseroja.
- 5) Kirjoittajan käyttäessä rikasta kieltä, mikään ei nouse esille.

Vaikka tämä on vain yhden bloggaajan yksilöllinen mielipide, on merkkejä, että monet karttavat sanapilviä juuri näistä syistä. Erityisesti turhat sanat ja sanojen lukuisat eri taivutusmuodot koetaan suurina ongelmina. Tämän pohjalta Analytica lähti rakentamaan työkalua, jolla se pystyisi poistamaan näitä ongelmia.

Sanapilven ehdoton vahvuus jo tälläkin hetkellä on sen kyky toimia erinomaisena tekstidatan visualisointityökaluna. (Magnusson, 2014). Tämän vuoksi, jos edellä mainitut ongelmat onnistutaan korjaamaan, on erittäin mahdollista saada työkalu myös suuren yleisön suosioon.



## 4.1 Ongelmien ratkaiseminen

Työkalua rakennettaessa Analystica löysi Turun yliopiston kehittämän työkalun aineiston lemmatisointiin. Lemmatisointi eli sanojen normalisoiminen tai perusmuotoistaminen tarkoittaa sitä, että aineistosta poimitaan kaikki sanat ja tehdään niistä perusmuotoisia. Tällöin esimerkiksi *kissan*, *kissalle*, *kissojen* muodot muutetaan, jolloin kaikista tulee *kissa*.

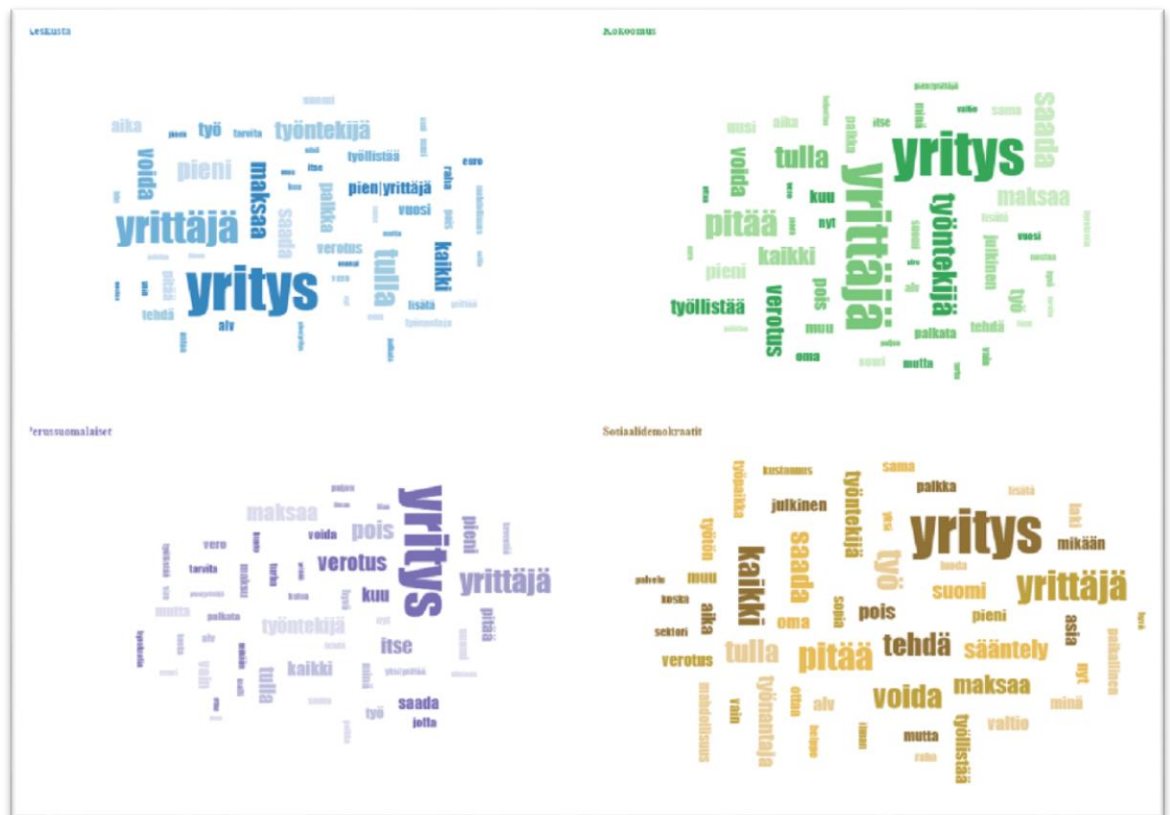
Analystica Oy:n lisäksi projektissa on vahvasti mukana eräs Åbo Akademin maisteri, joka tekee aiheesta tohturityötä. Yhteistyössä hänen kanssaan on työkaluun saatu luotua jo ylimääräisten sanojen poistomekanismi, johon käyttäjä pystyy halutessaan lisäämään poistettavia sanoja.

Tämän lisäksi kehitteillä on synonyymisanastojen luontimahdollisuus. Tämä toimisi vastaavalla tavalla. Synonyymisanasto sisältäisi automaattisesti tiettyjä sanoja, mutta käyttäjän olisi mahdollista halutessaan lisätä siihen uusia sanoja.

Kehitteillä on myös mahdollisuus tutkia yleisempiä sanapareja ja sanakombinaatioita. Tämä auttaisi käyttäjää ymmärtämään nopeammin, minkä sanojen yhteydessä kustakin aiheesta puhutaan.

Yksi tärkeimpiä uuden työkalun ominaisuuksia on mahdollisuus palata tekstiaineistoon sen jälkeen, kun on löytänyt jotain kiinnostavaa. Kuten kappaleessa 2.2 mainitaan, sisällönanalyysi, jota avainsanapilvet ovat, ei vielä itsessään kerro muuta kuin suuntaa-antavia tuloksia. Tämä ominaisuus mahdollistaa sen, että käyttäjä löytää ensin kiinnostavan aihealueen ja pääsee myös tutkimaan, mitä kyseisestä aihealueesta puhutaan tai kirjoitetaan.

## 4.2 Vertailumahdollisuudet



Kuva 2. Eri puolueiden vaalilupauksien vertailu.

Mainittujen viiden ongelman ratkaisemisen lisäksi ideana on kehittää kaksi erilaista tapaa tutkia aineistoja. Ensimmäinen tapa, jota on käytetty esimerkkikuva kakkosessa, olisi eri aineistojen tutkiminen yksittäisenä, mikä on tällä hetkellä yleisin tapa.

Toinen, hieman uudenlainen tapa olisi vertailla kahta tai useampaa eri aineistoa keskenään kahdella tavalla. Joko tutkia, mitä kussakin aineistossa nousee esille, mitä toisissa ei löydy lainkaan tai vaihtoehtoisesti tutkia, mitä kaikissa aineistoissa yhteisesti nousee esille.

### 4.3 Tiedon visualisointi

Tapio Nurminen (2013) nostaa artikkelissaan *Yhdeksän kysymystä tiedon visualisointiin* esille yhdeksän seikkaa, joihin tulisi pystyä vastaamaan, jotta datasta saataisiin mahdollisimman paljon irti.

Koska suunnitteilla olevan työkalun yksi osa on tiedon visualisointi, johon tarkoituksena on käyttää avainsanapilviä, pohdin näitä yhdeksää kysymystä toteutettavan projektin näkökulmasta.

#### *1. Mikä tavoite visualisoinnilla on?*

Avainsanapilvien tavoitteena on esittää selkeämmin kunkin aineiston eri aihealueet. Tällöin käyttäjä saa selkeän kuvan siitä, mitä hänen kannattaa tutkia tarkemmin.

#### *2. Ketkä ovat kohdeyleisö ja missä kontekstissa?*

Kohdeyleisönä ovat asiakkaat, joilla on isoja määriä avointa tekstimassaa. Työkalun tarkoituksena on helpottaa ja nopeuttaa tällaisten aineistojen tulkintaa.

#### *3. Miten palvelut käyttäjiä parhaiten?*

Työkalu säästää käyttäjänsä aikaa sekä tehostaa työntekoa.

#### *4. Mitä visualisoinnin tulisi sisältää?*

Visualisoinnin tulisi sisältää selkeä kuva niistä aihealueista, joita tekstissä käsitellään, sekä mahdollisuus palata aineistossa niihin kohtiin, missä kutakin aihetta käsitellään.

#### *5. Mitä tietoja visualisointiin tarvitaan?*

Visualisointiin riittää ainoastaan tiivistelmä aineiston aiheista. Kuten aikaisemmassa vastauksessa kerroin, tärkeintä on, että käyttäjä pääsee kätevästi takaisin aineistoon haluamaansa kohtaan.

### *6. Kuinka keräämme ja ylläpidämme tarvittavaa dataa?*

Analystica tarjoaa kehittyneitä tiedonkeruupalveluita, joiden tueksi tekstinanalysointityökalu on suunnitteilla. Lisäksi tekstin analysointityökalua voi käyttää minkä tahansa asiakkaan oman aineiston analysointiin.

### *7. Mitä lähestymistapaa projektissa käytämme?*

Pääasiallisena visualisointikeinona toiminevat avainsanapilvet.

### *8. Kuinka teemme visualisoinnista riittävän kiinnostavan mutta helppokäyttöisen?*

Helppokäyttöisyys muodostuu siitä, että aineistoon palaaminen avainsanapilven kautta tehdään mahdolliseksi. Visualisoinnin kiinnostavuus itsessään on taas riippuvainen aineiston kiinnostavuudesta. Mikäli aineistossa ei käsitellä mitään kiinnostavaa aihetta, ei visualisointi myöskään ole itsessään kiinnostava. Sanapilven taustalla on myös aina taulukko, jonka käyttäjä saa halutessaan näkyviin. Tämä saattaa myös helpottaa käyttöä.

### *9. Ymmärtääkö kohdeyleisö visualisoinnin?*

Avainsanapilvien peruseräite on sen verran selkeä, että sen ymmärtämiseen tarvitaan ainoastaan hieman loogista ajattelua. Tämän lisäksi avainsanapilvien rakenne pyritään selvittämään mahdollisimman tarkasti sekä työkalun nettisivuilla että Analystican omilla nettisivuilla.

## 5 TUTKIMUS

Projektin edetessä avoimen tekstin analysointityökalusta on keskusteltu muutamien eri tahojen, mm. HUS:n, Pirkanmaan sairaanhoitopiirin (PSHP) sekä Turun kaupungin edustajien, kanssa.

Yhteistä kaikissa keskusteluissa oli se, että avainsanapilvet, joiden on tarkoitus olla analysoinnin pääkeinona, herättivät aluksi negatiivisia reaktioita. Mutta kaikissa tapauksissa, kun pääsimme kertomaan, mitä työkalu pystyy aineistoille tekemään, mm. lemmatisointi, alkoi työkalu kiinnostaa suuresti.

Erittäin tärkeänä asiana kaikki heistä pitivät mahdollisuutta palata suoraan sanapilvestä aineistoon, kun löytää jotain kiinnostavaa.

Turun kaupungin edustajan kanssa keskustelu ohjautui siihen, miten kaupungin on tarkoitus 6aika-projektin mukaisesti avata eri aineistojaan yleiseen käyttöön. Tämä lisäisi mahdollisuuksia hyödyntää avoimen tekstin analysoinnin työkalua suuresti. Tällä hetkellä Turun kaupungin eri aineistoihin on vaikeaa tai mahdotonta päästä käsiksi.

### 5.1 Menetelmä

Markkinatutkimuksen pääasiallisena menetelmänä olivat puhelinhaastattelut. Puhelinhaastattelut valittiin sen vuoksi, että aihe ei ole täysin yksiselitteinen, eikä siitä siksi ole kovin helppo esittää tarkkoja kysymyksiä saamatta ensin vastauksia muutamiin tarkentaviin kysymyksiin. Puhelinhaastattelua voi siis ohjailta hieman haluamaansa suuntaan, kun saa ensin tietää hieman haastateltavan lähtökohdista.

## 5.2 Sähköpostikutsut

Päädyimme yhdessä toimeksiantajan kanssa lähettämään sähköpostiviestillä kutsun, jossa esiteltiin opinnäytetyön aihe sekä toivomus suorittaa noin 15 minuutin pituinen puhelinhaastattelu.

Lähetimme kutsun Analystican ohjelmiston avulla, mikä helpotti seurantaan siitä, ketkä olivat vastanneet ja mitä he olivat vastanneet. Tämän lisäksi järjestelmällä pystyimme antamaan kutsun vastaanottajille mahdollisimman helpon tavan vastata kutsuun.

Kutsu sisälsi neljä eri linkkiä, neljän mahdollisen tilanteen varalle.

1. He, jotka olivat kiinnostuneet osallistumaan puhelinhaastatteluun, pääsivät ensimmäistä linkkiä klikkaamalla ilmoittamaan, milloin heitä sopisi haastatella ja mistä numerosta heidät tavoittaa.
2. Heille, jotka eivät halunneet puhelinhaastattelua, mutta olivat valmiita osallistumaan muulla tavoin, annettiin mahdollisuus toista linkkiä klikkaamalla päästä nettikyselyyn, jossa selvitettiin samoja asioita kuin puhelinhaastattelussa oli tarkoitus selvittää.
3. Mikäli kutsu oli tullut sellaiselle henkilölle organisaatiossa, joka ei ollut tekemisissä suurien datamassojen kanssa, oli hänellä mahdollisuus siirtää kutsu kolmatta linkkiä klikkaamalla toiselle organisaation jäsenelle.
4. Neljännen linkin loimme niille, jotka eivät olleet lainkaan kiinnostuneita kyselystä. Tätä klikkaamalla he varmistivat, että heitä ei häiritäisi enää uudestaan asian tiimoilta.

Kutsussa ilmaistiin, että puhelinhaastattelu olisi toivottavin ratkaisu, mutta kaiken varalle luotiin myös nettikysely. Nettikyselyn tulosten analysointi ja luotettavuus tulisi arvioida huolellisesti, koska nettikyselyssä ei pysty varmistamaan, että vastaaja on ymmärtänyt asian oikein, eikä pääse ohjaamaan haastattelua haluamaansa suuntaan.

Kutsun linkkien avulla pyrittiin tarjoamaan jokaiselle vastaanottajalle tyydyttävä ratkaisu vastata. Saimme myös helposti tietoa niistä, joita asia ei lainkaan koskettanut tai kiinnostanut.

Sähköpostikutsuja lähetettiin yhteensä 171 kappaletta. Alkuperäinen luku oli 150 kappaletta, mutta 9 vastaajaa siirsi kutsun toiselle henkilölle organisaatiossa sekä 12 vastaanottajaa lisättiin vielä kesken tutkimuksen.

Tavoitteena oli kerätä 15–20 vastausta (noin 10 % kutsutuista), jotta saataisiin mahdollisimman tarkka kuva markkinatilanteesta.

Vastausaikaa annettiin noin kuukausi. Puolesta välissä tutkimusta lähetettiin muistutusviesti niille, jotka eivät olleet kutsuun reagoineet vielä millään tavoin.

### 5.3 Kohderyhmä

Kohderyhmäksi pyrimme kokoamaan kolme eri segmenttiä: työntekijämäärältään suuret yritykset, jäsenmäärältään suuret yhdistykset sekä markkinoiden suurimmat konsulttifirmat. Nämä kohderyhmät valittiin yhdessä toimeksiantajan kanssa sillä ajatuksella, että näillä yrityksillä olisi todennäköisesti käytössään suuria määriä datamassaa, jota he mahdollisesti jo analysoivat jollain tavoin tai ehkä toivovat, että löytäisivät työkalun analysointia varten.

Yrityksen lisäksi oli tärkeää löytää henkilö, jonka työnkuvaan liittyisi jonkinlainen analysointi. Tavoitteena oli löytää asiakaspalvelupäällikkö tai jokin sitä vastaava työntekijä yrityksestä. Joidenkin yritysten kohdalla päädyttiin mm. viestintäpäällikköön tai markkinointipäällikköön, koska muita ei löytynyt. Tässä vaiheessa ajatuksena oli, että vaikka viesti menisi väärälle henkilölle, voisi vastaanottaja siirtää sähköpostikutsun kohdan 3. mukaisesti kutsun toiselle henkilölle yrityksessä helposti yhden linkin avulla.

### 5.4 Puhelinhaastattelut

Puhelinhaastattelut suoritettiin aina vastaajan valitsemana ajankohtana. Haastattelun aluksi käytiin läpi vastaajan tiedot ja varmistettiin, että haastattelu-aika on vastaajalle sopiva. Haastattelut kestivät keskimäärin 15 minuuttia.

Haastattelut olivat rakenteeltaan kaikki samankaltaisia, mutta jokaisessa edettiin myös hieman vastaajan vastauksien mukaisesti, jotta haastattelusta saatiin kokoon mahdollisimman keskustelunomainen.

Haastattelun aluksi käytiin läpi, millainen työkalu oli suunnitteilla; Mikä sen tarkoitus oli ja millaisia ominaisuuksia se sisältää. Kun haastateltava oli saanut tarpeeksi hyvän käsityksen työkalusta, siirryttiin selvittämään haastateltavan ja hänen yrityksensä tilannetta.

Tarkoituksena oli saada selvillä, millaista dataa, jota työkalulla voisi analysoida, yritys kerää sekä mitä mahdollisia työkaluja tai menetelmiä he jo sillä hetkellä käyttävät. Mikäli yritys käytti muita vastaavia työkaluja, pyrittiin selvittämään, mitä nämä ovat sekä mitä huonoja ja hyviä puolia niissä on.

Haastattelun loppupuolella palattiin vielä Analystican valmistamaan työkaluun. Pyrittiin selvittämään, mikä ominaisuus tai mitkä ominaisuudet haastateltavalle olivat jääneet mieleen alun läpikäynnistä sekä mitä mahdollisia uusia ominaisuuksia haastateltava voisi kaivata vastaavanlaiselle työkalulle.

Viimeiseksi selvitettiin haastateltavan ja hänen yrityksensä halua osallistua maksuttomaan kokeiluun, jossa he antaisivat yrityksensä aineistoa Analysticalle, joka tuottaisi siitä ilmaisen koanalyysin. Tämän tarkoituksena oli löytää muutamia erilaisia aineistoja testiksi sekä antaa haastateltaville palkinnoksi ilmaisia työnäytteitä.

Kaikkiaan puhelinhaastatteluja suoritettiin kymmenen kappaletta, joista kaksi lopetettiin kohtuullisen nopeasti, kun yhdessä todettiin, että vastaajan ammatti-asema tai työtehtävä ei sopinut kohderyhmään. Tavoitteesta siis jäätii hieman tämän osalta.

## 5.5 Internetkyselylomake

Sähköpostikutsun toisen linkin kautta oli mahdollista siirtyä suoraan netissä sijaitsevaan kyselylomakkeeseen. Lomakkeen rakenne oli vastaava kuin puhelinhaastattelun rakenne.



Ongelmaksi alun alkaen koettiin, että laskeutumissivulla oli liikaa tekstiä. Tästä ei kuitenkaan päästy eroon, sillä yhdessä toimeksiantajan kanssa todettiin, että ilman tarpeeksi selkeää työkalun esittelyä alussa vastaaja ei välttämättä osaisi vastata oikealla tavalla kysymyksiin.

Kaikkiaan 18 vastaajaa päätti vastata nettikyselyyn. Näistä kuitenkin ainoastaan neljä vastaajaa oli oikeasti vastannut jotakin aiheeseen liittyvää kysymyksiin. Loput 14 olivat todennäköisesti luovuttaneet laskeutumissivun suuren tekstimäärän kohdalla, mikä jo aiemmin tiedettiin pullonkaulaksi.

## 6 TULOKSET

171 lähetettyyn sähköpostikutsuun saatiin 92 vastausta (54 %). 92 vastaajasta 58 (63 %) vastasi, ettei ole lainkaan kiinnostunut kyselystä. Analysointiin saakka päätyi 12 vastaajaa (8 puhelinhaastattelusta, 4 nettikyselystä), mikä on seitsemän prosenttia kokonaiskutsujen määrästä ja 13 prosenttia kaikista vastanneista.

Analysoitavien vastausten joukkoon kertyi erilaisista taustoista olevia henkilöitä. Tämä näkyi myös mm. tarpeissa ja datamäärissä, joissa oli selkeitä eroja. Tämänkin pienen otannan perusteella voi kuitenkin arvioida, että harvat yritykset vielä tälläkään hetkellä käyttävät vastaavia työkaluja aineistojensa analysointiin, vaikka tarvettakin olisi. Pääosin analysointi hoidetaan manuaalisesti.

Käytetyistä työkaluista yhtäkään ei mainittu käytettävän useammassa kuin yhdessä yrityksessä. Käytettyjä työkaluja olivat Wordle-sanapilvi, Webropolin palvelu, SPSS-palvelu, jonkinlaiset hakuraportit sekä yritysten kehittämät omat työkalut.

Lähes jokaisen työkalun kohdalla niiden käytännöllisyyttä kyseenalaistettiin jonkin verran. Manuaalisiin keinoihin luottavat yritykset kertoivat prosessin olevan työlästä, vaivalloista ja avaavan liikaa mahdollisuuksia subjektiiviselle tulkinnalle.

Eräs vastaajista oli tutustunut useamman toimittajan sanapilvipalveluihin ja antoikin rohkaisevan lausunnon siitä, että huonoimmat ovat surkeita ja parhaimmatkin ovat vain lähellä Analystican vastaavaa kehitteillä olevaa. Lisäksi lisenssihintaa per käyttäjä on parhaimmissa palveluissa erittäin korkea.

Analystican työkalun hyviä puolia selvitettäessä nousi esille perusmuotoistus (erityisesti suomen kielen osalta), vertailumahdollisuus, synonyymisanasto, sanaparit sekä mahdollisuus porautua sanapilvestä syvemmälle aineistoon. Erityisesti eri aineistojen vertailumahdollisuus nousi useasti esille ja oli selkeästi uusi asia vastaajien keskuudessa.

Selvitettäessä lisämahdollisuuksia työkalulle nousi esille mm. eri kielien käyttö perusmuotoistamiseen, yksinkertaisuus visualisoinnissa, mediassa liikkuvan tiedon analysointimahdollisuudet sekä niiden luokittelu ja tiivistelmien tekeminen alkuperäisestä tekstistä.

Osa näistä on tällä hetkellä hieman vaikeammin toteutettavia, ainakin niin, että se olisi kustannustehokasta. Eri kieliversioita perusmuotoistamiseen sen sijaan Analystica on jo löytänyt useampia. Visualisoinnin yksinkertaisuus on myös ollut yhtenä ajatuksena työkalussa, vaikka sitä ei varsinaisesti tässä tutkimuksessa tutkittu.

Analysointi tarpeet liittyivät useimmiten palautteisiin tai erilaisiin kyselyihin, mikä ei sinänsä ole yllättävää, sillä kohderyhmäksi pyrittiin hakemaan asiakaspalvelusekä markkinointialan ammattilaisia. Valitettavasti yhdeltä kolmesta kohderyhmästä, konsulttiyritykset, ei saatu vastauksia ollenkaan, joten niiden tarpeista ei ole vielä mitään käsitystä.

Maksuttomaan kokeiluun osallistumisesta kiinnostuneita olivat melkein kaikki vastanneet. Tämä ei varsinaisesti ollut tutkimuksen päätarkoitus, mutta antaa lisäarvoa suuresti. Nyt Analystica pääsee valitsemaan kiinnostavia testiaineistoja muutamalta tarjoajalta.

Analysoinnin johtopäätöksenä voidaan todeta, että tarvetta työkalulle on, mikäli työkalu on kustannustehokas. Myös kehitteillä olevissa ominaisuuksissa on osuttu oikeaan suuntaan. Työkalun ominaisuuksilla pystytään välttämään ongelmakohtia, jotka avoimia tekstejä analysoineet ovat kohdanneet.

## 6.1 Tutkimuksen luotettavuus

Tutkimuksen luotettavuutta pohdittaessa päällimmäisenä nousee esille tavoiteltua pienempi vastaajamäärä. 12 kunnollista vastausta on kuitenkin lähellä 15-20 haarukkaa, joka tavoitteeksi asetettiin. Toisaalta on syytä uskoa, että valikoituneet vastaajat edustivat kohderyhmää kohtuullisen hyvin, sillä heitä haastateltaessa sai selkeästi sellaisen kuvan, että he ymmärtävät, mistä puhutaan.

Kaiken kaikkiaan pitäisin tutkimusta luotettavana, joskin suppeana. On vaikeaa sanoa, olisiko suurempi vastaajamäärä tuottanut erilaisia vastauksia tai erilaisia johtopäätöksiä.

## 6.2 Pohdinta

Tämän opinnäytetyön tavoitteena oli selvittää Analystica Oy:n avoimen tekstin analysointityökalun tarvetta ja mahdollisuuksia markkinoilla sekä tutkia yleisimpiä ongelmia, joita vastaavanlaisilla työkaluilla esiintyy.

Tutkimus suoritettiin puhelinhaastatteluin sekä muutaman vastaajan osalta nettikyselyyn vastaamalla. Vaikka vastauksia saatiin hieman alle odotusarvon, voidaan tutkimusta pitää kohtuullisen onnistuneena.

Tutkimuksen avulla saatiin selville useampia työkalusta kiinnostuneita yrityksiä, heidän tarpeitaan analysoinnin osalta sekä arvokasta tietoa siitä, mitkä työkalun ominaisuudet kiinnostavat yrityksiä.

Päätavoitteena olleen, työkalun markkinamahdollisuuksien tutkiminen, lisäksi saatiin siis selville vinkkejä siitä, miten työkalua kannattaa markkinoida, eli mikä ihmisiä kiinnostaa, sekä saatiin kerättyä useita mahdollisia ehdokkaita työkalun maksuttomaan kokeiluun varten.

Tutkimuksen parhaana puolena voikin todeta, että se oli kustannustehokas tapa kerätä tietoa tutkimuksen päätavoitteen lisäksi myös muista asioista. Koska haastattelun toteuttaminen vaati työkalun toiminnan selvittämistä vastaajalle, on myös työkalun kohdemarkkinointi aloitettu samalla.

Tutkimuksen tulokset eivät poikenneet kovin paljon odotetusta. Päinvastoin, ne vain varmistivat jo aiemmin ajateltua tilannetta. Suurimpana yllätyksenä voidaan pitää vertailumahdollisuuden nouseminen esiin useiden vastaajien keskuudessa sekä sen herättämä suurehko kiinnostus.

Opinnäytetyön koko prosessia ajatellessa toimeksiantajan sekä opinnäytetyön tekijän yhteistyö toimi erinomaisesti. Molemmat olivat suurimman osan ajasta samoissa ajatuksissa, miten asioita lähestytään. Tätä helpotti tietenkin käynnissä ollut työharjoittelu toimeksiantajalla.

Yhteistyö kolmannen osapuolen eli Turun ammattikorkeakoulun kanssa taas oli hieman haastavaa välillä. Toimeksiantajan tarpeiden tyydyttäminen ei kaikissa tilanteissa vastannut ammattikorkeakoulun tarpeita. Tämän vuoksi projektin su-lava eteneminen tökki välillä hieman.

Lopulta voidaan kuitenkin todeta opinnäytetyön koko prosessin onnistuneen koh-tuullisesti. Tavoitteet saavutettiin kohtuullisesti ja aikataulukin oli lähes suunniteltu. Tutkimuksen ehkä syvin tarkoitus, antaa lisäarvoa toimeksiantajalle, onnistui mielestäni hyvin.

## LÄHTEET

6Aika 2015. viitattu 27.5.2015 <http://6aika.fi/>.

Hague, P. & Jackson P. 1999. Market research a guide to planning, methodology and evaluation, second edition. Lontoo: Kogan page limited

Kortesuo, K. 2013. Avainsanapilvi on usein vain kaunis kikkare. Viitattu 5.5.2015 <http://eioototta.fi/avainsanapilvi-on-usein-vain-kaunis-kikkare/>.

Nurminen T. 2013. Yhdeksän kysymystä tiedon visualisointiin. viitattu 26.6.2015 <http://www.floapps.com/yhdeksan-kysymysta-tiedon-visualisointiin/>.

Magnusson, C. 2014. Puolustuspuhe Sanapilvella. Viitattu 5.5.2015 <http://www.instanssi.fi/puolustuspuhe-sanapilvelle/>.

Salo I. 2013. Big data tiedon vallankumous. Jyväskylä: Docendo.

Salo I. 2014. Big data ja pilvipalvelut. Jyväskylä: Docendo.

Taloustutkimus Oy 2015. Kvalitatiivinen tutkimus. Viitattu 5.5.2015 [http://www.taloustutkimus.fi/tuotteet\\_ja\\_palvelut/tiedonkeruuratkaisut\\_ja\\_monitila/kvalitatiivinen\\_tutkimus/](http://www.taloustutkimus.fi/tuotteet_ja_palvelut/tiedonkeruuratkaisut_ja_monitila/kvalitatiivinen_tutkimus/).

Tuomi J. & Sarajärvi A. 2008. Laadullinen tutkimus ja sisällönanalyysi 11., uudistettu laitos. Helsinki: Tammi.