

Heta Pärssinen

Google Analyticsiin perustuvan datan esittäminen ja hyödyntäminen verkkosovelluksessa

Metropolia Ammattikorkeakoulu

Insinööri (AMK)

Tietotekniikan koulutusohjelma

Insinöörityö

11.10.2016

<p>Tekijä Otsikko</p> <p>Sivumäärä Aika</p>	<p>Heta Pärssinen Google Analyticsiin perustuvan datan esittäminen ja hyödyntäminen verkkosovelluksessa</p> <p>40 sivua 11.10.2016</p>
<p>Tutkinto</p>	<p>Insinööri (AMK)</p>
<p>Koulutusohjelma</p>	<p>Tietotekniikan koulutusohjelma</p>
<p>Suuntautumisvaihtoehto</p>	<p>Ohjelmistotekniikka</p>
<p>Ohjaaja</p>	<p>Lehtori Peter Hjort</p>
<p>Insinööriyön tarkoituksena oli tutkia Google Analyticsin mahdollisuuksia verkkosovelluksissa ja toteuttaa sen avulla palvelu, joka hyödyntää Google Analyticsistä saatavaa dataa. Palvelussa hyödynnetään esimerkiksi Google Analyticsistä saatavia sivuston lukukertoja, vierailijoiden lähdesivuja ja heidän käyttämiään laitteita.</p> <p>Google Analyticsille on olemassa useita rajapintoja, joita vertailtiin insinööriyössä. Käytettäväksi rajapinnaksi valittiin Embed API (API, application programmin interface, ohjelmointirajapinta). Toteutetussa palvelussa sivuston lukukerrat, vierailijoiden lähdesivut ja heidän käyttämänsä laitteet haettiin reaaliaikaisesti ja esitettiin käyttäjälle. Vanhempi data tallennettiin tietokantaan, josta dataa haettiin ja esitettiin käyttäjälle.</p> <p>Keskeisinä asioina olivat myös ohjelmointirajapintarajoitteet, datan omistajuus ja tallennus. Google asettaa ohjelmointirajapinnoille rajoitteet, jotka voivat olla käyttäjä-, IP- tai projekti-kohtaisia. Työssä testattiin Embed API:n rajoituksia. Googlen ohjelmointirajapinnasta saatavan datan tallentaminen ja myöhemmin käyttäminen on hyväksyttyä. Google Analyticsin tilin ylläpitäjä myös omistaa tiliin liittyvän Google Analytics -datan.</p> <p>Lopputuloksena syntyi toimiva palvelu, joka hyödyntää Google Analyticsin Embed API:n rajapintaa. Embed API oli sopivin ohjelmointirajapinta kyseiseen toteutukseen, ja ohjelmointirajapinnan dokumentaatio oli selkeä. Palvelussa käyttäjä voi nähdä sivuston lukukertoja ja tilastoja, joiden esityksessä hyödynnetään Embed API:a.</p> <p>Google Analyticsin ohjelmointirajapinnat ovat hyvin dokumentoituja ja luotettavia. Rajapintoja pystytään käyttämään eri tarkoituksiin, ja niiden avulla voidaan helpottaa Google Analyticsistä saatavan datan selkeyttä.</p>	
<p>Avainsanat</p>	<p>Google Analytics, verkkoanalytiikka, Google Analytics API, ohjelmointirajapinta</p>

Author Title	Heta Pärssinen Presentation and use of Google Analytics based data in web-application
Number of Pages Date	40 pages 11 October 2016
Degree	Bachelor of Engineering
Degree Programme	Information Technology
Specialisation option	Software Engineering
Instructor	Peter Hjort, Senior Lecturer
<p>This project focuses on researching the possibilities of Google Analytics in web-based applications and using this information to implement a service that uses data from Google Analytics. The service will use the data from Google Analytics, for example page views, sources of the visitors and the devices they use.</p> <p>Google Analytics has multiple application programming interfaces which are compared in this thesis. Embed API (an application programming interface) was chosen as a programming interface. Page views, sources of the visitors and the devices they use were requested real-time and showed to an end user. Older data was stored in the database. Later, data was fetched and showed to the user from the database.</p> <p>Application programming interface limits, data ownership and storing were fundamental issues in the project. Google sets limits for the interfaces which can be user, IP or project based. Embed API limits were tested in this project. Referring to data ownership and storing, it is allowed to save and use the data from Google programming interface. The administrator of Google Analytics account also owns the data from Google Analytics.</p> <p>The outcome of this project is a service that uses the interface of the Google Analytics Embed API. Embed API was the most convenient for this project and the documentation of the API interface was clear. The service provides page views of the site and statistics to the user. Embed API is used for providing this data.</p> <p>In conclusion, the Google Analytics programming interfaces are reliable and documented well. The programming interfaces can be used for different needs. The data from Google Analytics can be shown more clearly with Embed API.</p>	
Keywords	Google Analytics, web analytics, Google Analytics API, application programming interface

Sisällys

Lyhenteet ja käsitteet

1	Johdanto	1
2	Verkkoanalytiikka	2
2.1	Verkkoanalytiikan mittarit	3
2.2	Verkkoanalytiikka ja Big Data	4
3	Google Analytics -palvelu	5
3.1	Käyttöönotto	7
3.2	Tekninen toiminta	8
3.3	Käyttöliittymä	9
4	Google Analytics ohjelmointirajapinnat	10
4.1	Ohjelmointirajapintojen tarjoamat mahdollisuudet	11
4.2	Ohjelmointirajapintojen rajoitteet	17
5	Palvelun taustatutkimus ja suunnittelu	18
5.1	Tarvittava data ja ohjelmointirajapinnan valinta	19
5.2	Datan tarkkuus	19
5.3	Datan omistajuus ja oikeudet	23
5.4	Vastaavat palvelut	24
5.5	Riskit	24
5.6	Suunnittelun tulos	25
6	Työpöytänäköympäpalvelu	26
6.1	Palvelun sivunäkymät	29
6.2	Virheiden hallinta	34
7	Testaus ja ongelmat	36
8	Yhteenveto	38
	Lähteet	41

Lyhenteet ja käsitteet

AJAX	Asynchronous JavaScript and XML. Tekniikka, jonka avulla voidaan siirtää dataa niin, ettei koko verkkosivua tarvitse ladata uudelleen.
API	Application Programming Interface. Ohjelmointirajapinta, jossa eri ohjelmat voivat tehdä pyyntöjä ja vaihtaa dataa.
API rajapinta	Valmiiksi määritelty ohjelmointirajapinta palvelinpuolen ja kolmannen osapuolen välille.
Big Data	Suurten ja järjestämättömien tietomassojen kerääminen, säilyttäminen ja analysointi tietoteknisten ratkaisujen avulla.
DOM	Document Object Model. Kuvastaa HTML- tai XML-dokumentin rakennetta puuna. Sen rakennetta voi tutkia ja muokata esimerkiksi JavaScriptin avulla.
Eväste	Pieni tekstitiedosto, jonka selain tallentaa käyttäjän laitteelle. Evästeeseen tallennetaan eri tietoja käyttäjästä.
Flash	Multimedia-alusta esimerkiksi animaatioiden ja pelien tuotantoon.
GIF	Graphic Interchange Format. Häviötön bittikarttagrafiikan tallennusformaatti.
HTML	Hypertext Markup Language. Standardi kuvauskieli verkkosivujen ja -sovellusten luomiseen.
HTTP	Hypertext Transfer Protocol. Hypertekstin siirtoprotokolla, jota selain ja WWW-palvelin käyttävät tiedonsiirtoon.
IP-osoite	Internet Protocol. Tietokoneen yksilöivä numerosarja tietoverkossa.
JPG	Joint Picture Experts Group. Häviöllistä pakkausta käyttävä bittikarttagrafiikan tallennusformaatti.

Java	Oliopohjainen ohjelmointikieli.
JavaScript	Verkkosovelluksissa käytettävä dynaaminen komentosarjakieli, jonka avulla voi lisätä sivuille dynaamista toiminnallisuutta.
JSON	JavaScript Object Notation. Yksinkertainen avoimen standardin tiedostomuoto tiedonvälitykseen.
Kirjasto	Tässä työssä kirjastolla tarkoitetaan ohjelmointikirjastoa, joka on kokoelma valmiita aliohjelmia. Kirjastot helpottavat ohjelmointia, jolloin aliohjelmia ei tarvitse kirjoittaa aina uudelleen.
LinkedIn	Verkostoitumispalvelu, johon käyttäjät voivat ladata esimerkiksi ansioluettelonsa ja harrastuksensa ja saada suosituksia sekä suositella muita.
PNG	Portable Network Graphics. Häviötön bittikarttagrafiikan tallennusformaatti.
Python	Monipuolinen, suosittu ohjelmointikieli.
XML	Extensible Markup Language. Merkintäkieli, jolla tiedon merkitys on kuvattavissa tiedon sekaan.

1 Johdanto

Vuoden 2000 tienoilla ihmiset alkoivat ymmärtää internetin mahdollisuudet tiedon löytämiseksi. Käyttäjäkunta kasvoi nopeasti, ja uusia sivustoja syntyi. Organisaatiot halusivat tarjota relevanttia sisältöä kävijäkunnalleen. Mitkä asiat ovat olennaisia millekin yritykselle? Onko olennaista kääntää sivusto kymmenelle kielelle, hyväksyä viisi eri valuuttaa ja optimoida sivusto viisi vuotta vanhalle Internet Explorer -versiolle? Verkkoanalytiikan avulla pyritään antamaan vastaukset näihin kysymyksiin. Teoksessa *Advanced Web Metrics with Google Analytics* Brian Clifton kuvaa verkkoanalytiikkatyökalujen mahdollisuuksia seuraavasti: ne mahdollistavat verkkokokemuksen tutkimisen ja sen parantamisen [1, s. 23].

Verkkoanalytiikka on datan mittaamista, keräämistä, analysointia ja raportointia. Sen avulla tavoitellaan internetkäyttäytymisen ymmärtämistä ja optimointia. Datalla tarkoitetaan käyttäjien tekemiä toimintoja, kuten navigointia sivustolla, sivustolla vietettyä aikaa ja klikkaustapahtumia. Verkkoanalytiikkatekniikat määritellään usein sisäiseksi ja ulkoiseksi (*on-site and off-site*) verkkoanalytiikaksi. Käyttäjän toiminnot ovat osa sisäistä verkkoanalytiikkaa. Ulkoiseksi verkkoanalytiikaksi määritellään usein kolmansien osapuolten tarjoamat palvelut kuten käyttäjäkyselyt, markkinointiraportit tai kilpailijoiden vertailu. Molemmat näistä tähtäävät verkkoanalytiikan perusperiaatteeseen: verkkokokemuksen tutkimiseen ja parantamiseen. [2, s. 7674.]

Verkkoanalytiikka on osa usean yrityksen toimintaa. Se on usein pakollinen mittari yrityksille, joiden toiminta perustuu verkkosovellukseen. Verkkoanalytiikka on iso osa-alue, jolla voi olla vaikutusta markkinointiin, myyntiin ja teknologiakehitykseen sekä yleisesti yrityksen kehitykseen. Verkkoanalytiikan avulla mitataan ja tarkastellaan käyttäjien sivustoilla suorittamia toimenpiteitä.

Insinööriyön tarkoituksena on selvittää, miten verkkoanalytiikkaan perustuvaa dataa voidaan hyödyntää ja minkälaista dataa Google Analyticsistä on saatavilla sekä toteuttaa Google Analytics -ohjelmointirajapinnan avulla luotu työpöytänäköymä yrityksille. Työpöytänäköymässä on tarkoitus näyttää yritysten jättämiin ilmoituksiin liittyvää dataa, kuten lukukertoja kuukausi- ja päivätasolla.

Palvelu takaa käyttäjille helpomman mahdollisuuden nähdä työpaikkailmoituksiinsa liittyvää dataa. Palvelussa on tärkeää käyttäjäystävällisyys, joten huomiota kiinnitetään suorituskykyyn ja ulkoasussa mukautuvaan suunnitteluun (*responsive design*).

Insinööriyön tavoitteena on suunnitella ja toteuttaa toimiva palvelu, jota käyttäjät pystyvät käyttämään. Toteutuksen yhteydessä sovellusta tulee testata. Työn tavoite on myös selvittää yleisellä tasolla, miksi verkkoanalytiikka on tärkeää ja miten se vaikuttaa yritystoimintaan sekä miten Google Analytics toimii ja tukee tätä.

2 Verkkoanalytiikka

Ihmiset ovat olleet kiinnostuneita analytiikasta internetin alkuaajoista lähtien. Kävijälaskurit vilahtelivat useilla internetsivuilla, ja joissakin palveluissa oli myös mahdollista nähdä, ketkä olivat vierailleet sivuilla. Internetsivut hyödyntävät tätä ja käytössä voi olla maksullisia palveluja, kuten ”laajennettu vieraslista” ja ”pieni vieraslista” [3].

Verkkoanalytiikka, yksi digimarkkinoinnin vakiintuneista tekijöistä, kerää dataa verkosta. ”Jos et voi mitata sitä, et voi parantaa sitä” on yksi sitaateista, joita voidaan pitää verkkoanalytiikan tarkoituksena [1, s. 1; 4, s. 13]. Verkkoanalytiikan avulla voidaan mitata esimerkiksi kampanjoiden ja hakukoneoptimoinnin tehokkuutta ja sitä kuinka hyvin käyttäjät pystytään sitomaan sivustolle. Hakukoneoptimointi tarkoittaa verkkosivuston näkyvyyden parantamista hakukoneille eli käytännössä sitä, kuinka korkealle verkkosivusto sijoittuu esimerkiksi Googlen hakutuloksissa tietyillä hakusanoilla. Jos verkkosivusto on tärkeä osa liiketoimintaa, on myös analytiikka ja sivuston käytön mittaus tärkeä osa sitä. Toiselta kannalta analytiikkaa voidaan käyttää liiketoiminnassa myytäessä asiakkaille palveluita. Asiakkaille voidaan tarjota palveluita, joissa esimerkiksi luvataan tietty määrä lukukertoja heidän ilmoitukselleen. Tämän perustana käytetään usein verkkoanalytiikkaan perustuvia mittareita.

Liiketoiminta voi turvautua analytiikan lukuihin, ja joskus yrityksen johto saattaa olettaa analytiikkojen rakentavan mallit, jotka kasvattavat yritystoimintaa tai johtavat yritystä oikeaan suuntaan. Teoksessa Business Analytics Rahul Saxena ja Anand Srinivasan kuvaavat, että johtotiimin tulee toimia yhteen niin analytiikkatiimin kuin IT-tiiminkin kanssa, jolloin tiimit tukevat toisiaan ja analytiikkaa voidaan hyödyntää parhaalla mahdollisella tavalla [5 s. 4].

IT-tiimin kannalta analytiikkaa voidaan hyödyntää esimerkiksi verkkosuunnittelussa. Analytiikasta voidaan tarkastella esimerkiksi, millä selaimilla ja laitteilla sivustolla tai palvelussa yleensä vierailaan. Tietoa voidaan käyttää esimerkiksi päätettäessä, mitä selaimia sivuston tulee tukea tai kuinka mobiiliystävällinen sivuston halutaan olevan.

2.1 Verkkoanalytiikan mittarit

Verkkoanalytiikan mittareilla tarkoitetaan laskennallisia tilastotietoarvoja, jotka kuvaavat sivuston tapahtumia tai trendejä. KPI-mittarit (*key performance indicator*) ovat tehokkaan mittaamisen kannalta tärkeimmät kvantitatiiviset tunnusluvut. Teoksessa *Web Analytics 2.0* [6, s. 37] Avinash Kaushik kuvaa kahdeksan kriittistä verkkoanalytiikan mittaria:

1. Käynnit (*visits*)

Käynnit tarkoittavat teknisesti sessioita. Session aikana käyttäjä saapuu sivustolle, navigoi sivuston sisällä ja lopulta poistuu sivustolta. Tämä tapahtuma edustaa yhtä käyntiä.

2. Uniikit kävijät (*unique visitors*)

Uniikit kävijät lasketaan myös sessioiden perusteella, mutta erotuksena kävijän id tunnistetaan, jolloin jos sama kävijä saapuu sivustolle, hänet voidaan tunnistaa ja vain yksi käynti tallentaa.

3. Aika sivulla ja aika sivustolla (*time on page and time on site*)

Aika sivulla edustaa jokaisella sivustolla vietettyä aikaa. Aika sivustolla edustaa käyttäjän koko sivustolla viettämää aikaa. Tämän laskentaan käytetään myös sessioita.

4. Välitön poistumisprosentti (*bounce rate*)

Välitön poistumisprosentti kuvaa, kuinka monta henkilöä poistui sivulta siirtymättä eteenpäin eli klikkaamatta mitään linkkiä sivulla. Esimerkiksi välitön poistumisprosentti 78 kuvaa, että 78 % kävijöistä poistui sivulta klikkaamatta mitään linkkiä ja 22 % klikkasi jotakin linkkiä.

5. Poistumisprosentti (*exit rate*)

Poistumisprosentti kuvastaa, kuinka monta henkilöä poistui sivulta sen ollessa viimeinen sivu käyttäjän sessiossa. Jos käyttäjän halutaan esimerkiksi suorittavan tietyt toiminnot sivulla, voidaan tutkia viimeisen sivun poistumisprosenttia.

6. Konversioaste (*conversion rate*)

Konversioaste tarkoittaa lopputavoitteen suorittajia jaettuna kävijöillä. Google Analyticsissä voidaan asettaa eri tavoitteita (*goals*). Tavoitteelle voidaan asettaa aloitussivu ja lopetussivu. Jos esimerkiksi verkkokaupassa käyttäjä tarkastelee jotakin tiettyä kenkää, voidaan tavoitesivuksi asettaa ostotapahtuma ja aloitussivuksi tuotteen sivu. Jos kolme käyttäjää tarkastelee kenkää ja yksi suorittaa ostotapahtuman loppuun, on konversioaste 33,3 %.

7. Sitoutuminen (*engagement*).

Sitoutumista on haastavaa mitata, koska sitoutuminen on mittareista kvalitatiivisin. Jotkin analytiikkatyökalut pystyvät esimerkiksi mittaamaan, palaako kävijä sivustolle, mutta on vaikea määrittää, palaako kävijä sivustolle hyvän sisällön takia vai muista syistä, kuten jos käyttäjä joutuu huonon internetyhteyden vuoksi lataamaan sivua uudelleen.

2.2 Verkoanalytiikka ja Big Data

Nykypäivänä dataa kerätään lähes jokaisesta käyttäjien tekemästä tapahtumasta: verkkokaupassa vierailu, mainoksen klikkaus, luottokorttiosoite, osoitteenmuutos. Nämä ovat vain osa mahdollisista tapahtumista. Dataa myös tallennetaan huomattavasti enemmän kuin ennen: videoita, kuvia, lokitiedostoja. Näitä hyvin suuria järjestelemättömiä data-massoja kutsutaan Big Dataksi. Big Data muuttaa yritysten toimintamalleja ja markkinointitekniikoita. Big Datalle ei ole olemassa yhtä tarkkaa määritelmää, mutta usein sitä kuvattaessa käytetään esimerkiksi sanoja volyyymi (*volume*), vauhti (*velocity*) ja vaihtelevuus (*variety*) [7]. Tämä tarkoittaa käytännössä, että dataa on paljon, sen määrä kasvaa nopeasti ja se on yhä monimuotoisempaa eikä sillä ole selkeää rakennetta.

Dataa voidaan kohdentaa käyttäjäkohtaisesti aivan eri tavalla kuin ennen. Tämän kaiken perustana ja taustalla on verkoanalytiikka ja käyttäjien toimintojen seuranta. Eniten se vaikuttaa verkkokauppojen toimintaan, koska käyttäjän jokainen toimenpide ja tarkkaillut

tuotteet voidaan seurata. Kuitenkin kaikilla yrityksillä, jotka toimivat verkkosovellusten parissa, on mahdollisuus hyödyntää tätä käyttäjästä keräytävää dataa. [8, s. 161.]

Verkkoanalytiikan avulla on mahdollista analysoida sivustojen käyttäjäkunta ja tarjota sille kohdennettua mainontaa tai dataa. Verkkoanalytiikan avulla kerättyä dataa voidaan hyödyntää eri osa-alueilla yrityksen toiminnassa:

- Tuotekehitys: Tuotteen käyttäjien ymmärrys.
- Markkinointi: Mitkä markkinointikampanjat ovat tehokkaimpia.
- Hinnoittelu: Hintojen optimointi ”viimeisille” tuotteille.
- Myynti: Resurssien luokittelu ja toimeksiantojen mallinnus.
- Henkilöstöosasto: Tehokkaimpien työntekijöiden piirteiden ja käytöksen identifiointi. [9, s. 31.]

Data on väline, jota käytetään nykypäivänä usealla eri alueella. Yritysten on pysyttävä mukana muuttuvassa maailmassa, kun myös käyttäjät vaativat yhä enemmän. Data vaikuttaa tulevaisuudessa yhä enemmän yritysten liiketoimintoihin.

Big datan vaikutus liiketoimintaan on mullistava. Datasta tulee keskeisin kestävä kilpailuedun lähde ja sen varaan rakennetuista innovaatioista yrityksen tärkein resurssi. Ne, jotka heräävät datan tärkeyteen nopeimmin ja suurimmalla intensiteetillä, tulevat parantamaan suhteellista kilpailuasemaansa lähitulevaisuudessa nopeimmin. [10, s. 135.]

3 Google Analytics -palvelu

Maaliskuussa 2005 Google osti Urchin-yrityksen, joka oli luonut tuotteen, josta tuli Google Analytics -palvelun perusta. Marraskuussa 2005 Google julkaisi Analytics-palvelun, joka auttaa mittaamaan verkkosivujen ja markkinointikampanjoiden vaikutusta.

Google Analytics on nykyään suosituin analytiikkatyökalu. Sitä käytetään miljoonilla internetsivuilla datan mittaamiseen ja analysointiin. Sen avulla mitattava data on peräisin käyttäjistä, jotka ovat vierailleet sivustolla. Google Analyticsista saadaan esimerkiksi seuraavaa dataa:

- päivittäiset vierailijat
- keskimääräinen konversio
- eniten vieraillut sivut
- keskimääräinen sivustolla vierailuaika ja kuinka nopeasti vierailijat palaavat edelliselle sivulle
- vierailijoiden maantieteellinen sijainti
- käyttäjien siirtyminen palvelun muille sivuille. [1, s. 8.]

Konversio tarkoittaa loppuun suoritettua valmiiksi määriteltyä toimintoa internetsivulla, esimerkiksi ostotapahtuma verkkokaupassa. Edellä määriteltyä dataa voidaan käyttää esimerkiksi markkinointikampanjoiden analysointiin tai muuhun yrityksen tarpeisiin. Dataa voidaan käyttää hyväksi niin markkinoinnissa kuin myynnissä.

Marketing Land -sivuston mukaan 30–50 miljoonaa sivustoa käyttää Google Analyticsiä [11]. Google Analytics ei itse paljasta sivuston tarkkaa käyttäjämäärää. Riippuen lähteestä luvut vaihtelevat kuitenkin 30 ja 50 miljoonan käyttäjän välillä. Internetsivuilla käytettyjä ohjelmistoja seuraavan sivun, BuiltWithin, mukaan noin 30 miljoonaa käyttäjää käyttää Google Analyticsiä. Tämä on BuiltWithin mukaan noin 8,1 % koko internetin sivustoista. Koko sivustomäärä kattaa noin 363 miljoonaa sivustoa. Kuitenkin sivustoista, jotka BuiltWith määrittelee suosituimmaksi liikenteen perusteella, noin 69,7 % käyttää Google Analyticsiä. Tämä kattaa noin miljoona sivustoa. Kokonaiskattavuusprosentit ovat esitettynä kuvassa 1.

Coverage Totals	
Quantcast Top 10k 6,715 of 10,000	67.2%
Quantcast Top 100k 76,043 of 100,000	76%
BuiltWith Top Million 659,623 of 946,228	69.7%
Entire Internet 29,616,540 of 363,403,267	8.1%

Kuva 1. Google Analyticsin käyttö builtwith.com -sivuston mukaan [12].

Korkeimman käyttäjämäärän analysoi SimilarTech, noin 47 miljoonaa käyttäjää [13]. Samalla sivusto arvioi miljoonasta suosituimmista sivustoista 57,9 %:n käyttävän Google Analyticsiä ja koko internetin sivustoista 23,7 %:n. On hyvin vaikeaa arvioida Google Analyticsin tarkkaa ja oikeaa käyttäjämäärää, koska sivustot käyttävät eri analysointityökaluja ja menetelmiä. Ennen kuin Google julkistaa luvut, on käyttäjämäärästä olemassa vain arvioita.

3.1 Käyttöönotto

Google Analytics -tilin luominen on ilmaista. Rekisteröityminen tehdään omalla Google-tilillä. Rekisteröitymisen yhteydessä tulee valita seurattavaksi sivuksi joko verkkosivusto tai mobiilisovellus. Kaikki verkkosivustojen seurannat tapahtuvat nykyään Universal Analyticsin kautta. Analytics oli ennen jaettuna kahteen eri palveluun: ilmaiseen Google Analyticsiin ja maksulliseen Universal Analyticsiin. Universal Analytics sisälsi erityisominaisuuksia, jotka kaikki kuuluvat nykyään perus-Analytics-tiliin.

Google Analyticsin rekisteröitymislomakkeella kysytään sivuston perustiedot, minkä jälkeen käyttäjälle määritellään seurantatunnus ja sivuston seurantakoodi. Seurantakoodi on lyhyt JavaScript-koodi, esimerkki on esitettyinä kuvassa 2. Seurantakoodi tulee asettaa jokaiselle sivustolle, jonka analytiikkaa halutaan seurata.

```

<script>
(function(i,s,o,g,r,a,m){i['GoogleAnalyticsObject']=r;i[r]=i[r]||function(){
(i[r].q=i[r].q||[]).push(arguments)},i[r].l=1*new Date();a=s.createElement(o),
m=s.getElementsByTagName(o)[0];a.async=1;a.src=g;m.parentNode.insertBefore(a,m)
})(window,document,'script','https://www.google-analytics.com/analytics.js','ga');

ga('create', 'UA-82794603-1', 'auto');
ga('send', 'pageview');

</script>

```

Kuva 2. Google Analyticsin seurantakoodi.

Google Analytics -tiliin on mahdollista liittää useita käyttäjiä. Tiliin on mahdollista liittää useita sovelluksia ja tilin alle luoda useita seurantatunnuksia. Seurantatunnuksien alle voidaan luoda eri näkymiä, joiden avulla sivustojen hallintaa voidaan yksinkertaistaa ja selkeyttää.

3.2 Tekninen toiminta

Kun käyttäjä saapuu sivulle, jolle Analytics-seurantakoodi on asetettu, sivustolla oleva Analytics JavaScript -koodi viittaa JavaScript-tiedostoon, joka suorittaa seurantaoperaation Analyticsille. Analytics käyttää kolmea eri lähdettä seuratakseen käyttäjää. Nämä lähteet ovat

1. käyttäjän HTTP-pyyntö

Käyttäjän HTTP-pyyntö mille tahansa internetsivustolle sisältää tiedot esimerkiksi käyttäjän selaimesta, kielestä ja käytetystä laitteesta.

2. selain- ja laitetiedot DOM:sta

DOM:sta luettavat selain- ja laitetiedot voivat sisältää tarkempaa tietoa esimerkiksi Java- ja Flash-tuesta tai näytön resoluutiosta. DOM tarkoittaa tässä tapauksessa HTML-sivun rakennetta ja sen elementtejä.

3. evästeet

Analytics lukee ja asettaa myös ensimmäisen osapuolen evästeitä (*first-party cookies*). Ensimmäisen osapuolen eväste asetetaan suoraan julkaisijan eli sivuston puolesta. Tämä eväste tallennetaan vierailijan selaimeen sen sivuston puolesta, jolla käyttäjä vierailee. Tällöin informaatio vaihtuu käyttäjän ja sivuston välillä, ja tiedot käyttäjän toiminnoista siirtyvät suoraan Google Analyticsille. Evästeiden avulla Analytics voi lukea käyttäjän session ja esimerkiksi mainoskampanjatiedot. Analytics asettaa useita eri evästeitä ja käyttää niitä eri tarkoituksiin. Esimerkkinä Analytics asettaa evästeen nimeltään ”__utma”. Tämä eväste on voimassa 2 vuotta asetuksesta tai sen päivityksestä. Analytics käyttää tätä evästettä erottaakseen käyttäjät ja näiden sessiot. Eväste luodaan tai päivitetään, kun JavaScript-kirjasto on ladattu, toisin sanoen, kun käyttäjä on ladannut sivun. Tällöin evästeen tiedot lähetetään myös Analyticsille. [14; 1, s. 22]

Tiedot lähetetään takaisin Google Analytics -palvelimille GIF-kuvapyynnöllä, johon Analytics yhdistää listan parametreja käyttäjästä. Esimerkki GIF-pyyntöstä on esitettyä kuvassa 3.

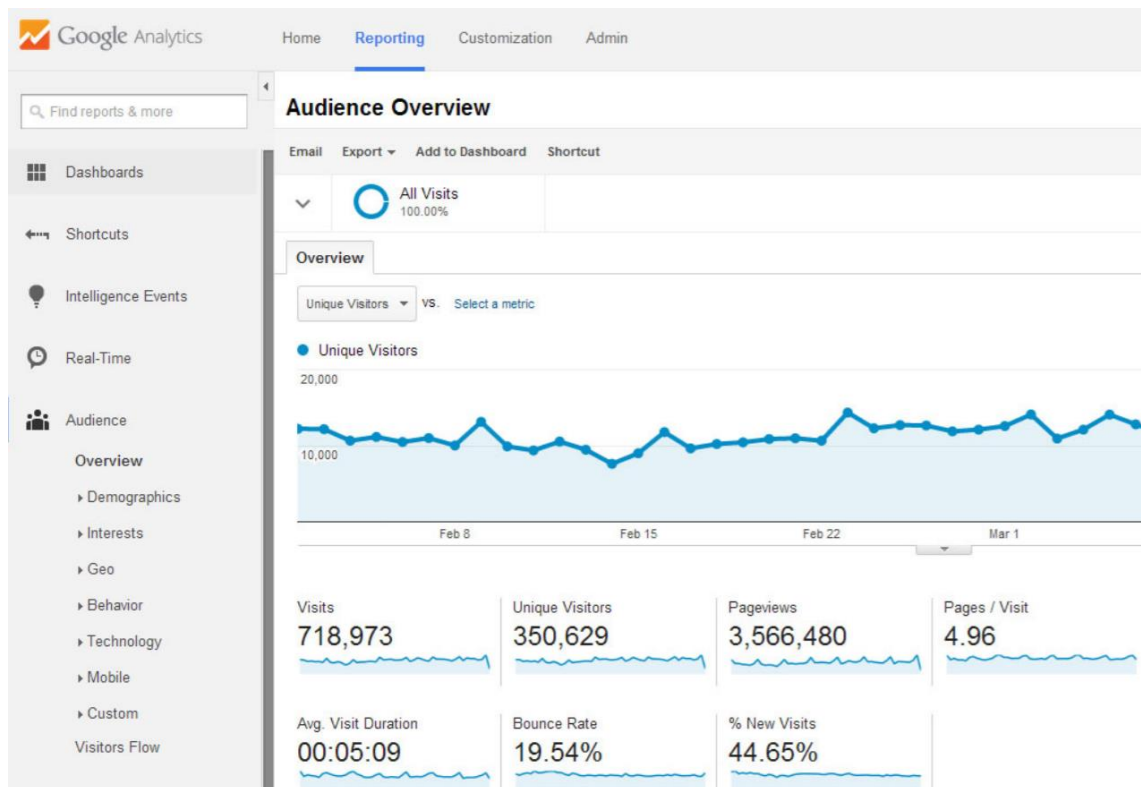
```
http://www.google-analytics.com/__utm.gif?utm-
mwv=4&utmhn=example.com&utmcs=ISO-8859-1&utm-
msr=1280x1024&utmcs=32-bit&utmul=en-us&utmje=1&utm-
mfl=9.0%20%20r115&utmcn=1&utmdt=GATC012%20setting%20variables&utm-
mhid=2059107202&utmr=0&utmp=/auto/GATC012.html?utm_source=www.gatc0
12.org&utm_campaign=cam-
paign+gatc012&utm_term=keywords+gatc012&utm_content=con-
tent+gatc012&utm_medium=medium+gatc012&utmacc=UA-30138-1&utm-
mcc=__utma%3D97315849.1774621898.1207701397.1207701397.1207701397.1
%3B...
```

Kuva 3. Esimerkki Google Analytics -palvelimelle lähetettävästä GIF-kuvapyynnöstä [14].

Analytics käyttää GIF-pyyntöä sen nopeuden ja pienen koon vuoksi. Verrattuna esimerkiksi PNG- ja JPG-kuvapyyntöihin GIF-kuvat vaativat vähiten tilaa bittiä kohden [15].

3.3 Käyttöliittymä

Google Analytics tarjoaa käyttäjälle työpöytä näkymän (*dashboard*). Kuvassa 4 on esitettyä Google Analyticsin Yleisö-työpöytä näkymä (*audience*). Näkymästä voi seurata esimerkiksi kävijöitä, uniikkeja kävijöitä, katselukertoja, sijaintitietoja ja laitetietoja.



Kuva 4. Google Analyticsin työpöytä näkymä.

Näkymissä sivuja on mahdollista tarkastella URL-osoitteiden perusteella. Google Analytics tarjoaa myös Käyttäytymisen näkymän (*behavior*). Näkymässä on yksityiskohtaisempaa tietoa käyttäjien käyttäytymisestä. Sivustolla on nähtävissä esimerkiksi kävijät, uniikit kävijät, keskimääräinen aika sivulla, välitön poistumisprosentti ja poistumisprosentti. Muita näkymiä Google Analyticsissa ovat konversiot (*conversions*), hankinta (*acquisition*) ja reaaliaikainen (*real-time*). Konversiot-näkymässä voidaan tarkastella asetettujen tavoitteiden toteutumista. Hankinta-näkymä antaa tarkempaa tietoa sivustoista, joista vierailijat saapuvat sivustolle. Reaaliaikainen näkymä taas antaa reaaliaikaista tietoa sivuston tapahtumista, esimerkiksi sivustolla olevat kävijät.

4 Google Analytics -ohjelmointirajapinnat

Ohjelmointirajapinta (*application programming interface - API*) tarjoaa rajapinnan jonkin kahden sovelluksen välille. Kaksi sovellusta voivat vaihtaa esimerkiksi tiedostoja keskenään ohjelmointirajapinnan avulla. Google Analytics tarjoaa erilaisia ohjelmointirajapintoja kehittäjille datan hyödyntämiseen. Analytics on kehittänyt seuraavat rajapinnat:

- Core Reporting API

- Embed API
- Multi-Channel Funnels Reporting API
- Real Time Reporting API
- Metadata API
- Management API
- Provisioning API.

Core Reporting API toimii perustana ja Embed API, Multi-Channel Funnels Reporting API ja Real Time Reporting API käyttävät sitä.

4.1 Ohjelmointirajapintojen tarjoamat mahdollisuudet

Core reporting API on Analyticsin ”kehittynein” ohjelmointirajapinta, ja se tarjoaa pääsyn Analyticsissa raportoituun dataan. Reporting API:n avulla voi

- rakentaa työpöytiä ja esittää Google Analyticsin dataa
- automatisoida vaikeampia raportteja ja säästää aikaa
- yhdistää Google Analytics datan muihin sovelluksiin.

Tämänhetkinen versio ohjelmointirajapinnasta on V4, ja se mahdollistaa pääsyn seuraaviin ominaisuuksiin:

- ulottuvuus (*dimension*) ja metriset (*metric*) tiedot
- useat data-alueet (*data range*) samassa kyselyssä
- kohortti ja elinikäiset arvot

- useat segmentit samassa kyselyssä. [16.]

Ulottuvuudet ovat datan attribuutteja, esimerkiksi "kaupunki" tai "sivu URL" voi olla ulottuvuus. Metrinen arvo on numeerinen arvo, esimerkiksi "sessiot" on metrinen arvo, joka kertoo sessioiden kokonaismäärän. Näiden arvojen perusteella voidaan luoda yhdistelmiä, ja useita arvoja voidaan käyttää samassa kyselyssä. Google Analytics "Dimensions & Metrics Explorer" -sivustolla [17] ovat nähtävissä kaikki mahdolliset ulottuvuus ja metriset tiedot, jotka ovat käytettävissä Core Reporting API:n kautta. Ohjelmointirajapinnan toimintaa voi testata Googlen luomalla "Query Explorer" -sivustolla [18].

Core Reporting API:iin luoduissa kyselyissä pakollisia parametreja ovat id, alkupäivämäärä, loppupäivämäärä ja metrinenarvo. Vapaaehtoisia arvoja ovat ulottuvuus (dimension), järjestys (sort), suodatin (filter), segmentti (segment), näytetaso (samplingLevel), tyhjien rivien mukaanotto (include-empty-rows), alkuindeksi (start-index) ja maksimitulokset (max-results). Suluissa oleva sana on muoto, jota kyselyissä tulee käyttää. Taulukossa 1 ovat kuvattuna kaikki mahdolliset kyselyyn liitettävät arvot.

Taulukko 1. Google Analytics Core Reporting API -kyselyyn liitettävien arvojen selitykset.

Nimi	Arvo	Pakollinen	Kuvaus
ids	string	kyllä	Lista uniikeista profiili ID:eistä Analyticsin datan hakuun.
start-date	string	kyllä	Kyselyyn mukaan otettavan datan alkupäivämäärä, joka voi olla muotoa YYYY-MM-DD tai today, yesterday tai NdaysAgo.
end-date	string	kyllä	Kyselyyn mukaan otettavan datan loppupäivämäärä, joka voi olla muotoa YYYY-MM-DD tai today, yesterday tai NdaysAgo.
metrics	string	kyllä	Metrics on kerätty laskettava arvo käyttäjäaktiivisuudesta sivustolla, kuten klikkaukset tai sivunäytöt, enintään 10 arvoa yhdessä kyselyssä.
dimensions	string	ei	Dimension arvo rajaa metrics-arvot jonkin kriteerin mukaan, esimerkiksi selaimen tai kaupungin. Enimmäismäärä on 7 dimensiota kyselyssä.
sort	string	ei	Lista arvoista, joilla halutaan järjestää tulokset. Arvot voivat olla dimension- tai metric-arvoja.
filters	string	ei	Suodattimet, jotka rajaavat kyselystä palautuvaa dataa. Arvot voivat olla dimension- tai metric-arvoja.
samplingLevel	string	ei	Haluttu otantalinja, sallitut arvot ovat DEFAULT, FASTER, HIGHER_PRECISION. Vaikuttaa kyselyn tarkkuuteen ja nopeuteen.
start-index	integer	ei	Kertoo ensimmäisen palautettavan rivin kyselystä.
max-results	integer	ei	Kertoo enimmäisrivimäärän, joka palautetaan kyselystä.

Esimerkiksi jos halutaan tarkastella, millä laitteilla uudet käyttäjät käyttävät sivustoa, voidaan laatia kysely, jossa metrisenä arvona käytetään uusia käyttäjiä ja ulottuvuutena laitekategoriaa. Esimerkkikysely ”Google Explorer” -sivustolle on esitettyä kuvassa 5.

Kuva 5. Kysely ”Google Explorer” -sivustolla Core Reporting API:iin.

Tuloksessa kuvassa 6 on eriteltyä pöytä-, mobiili- ja tabletilaitteet sekä niiden kävijämäärät 30 päivän ajalta, kuten kyselyssä oli asetettu.

Device Category	New Users
desktop	110
mobile	113
tablet	23

Kuva 6. Suoritetun kyselyn esimerkivastaus.

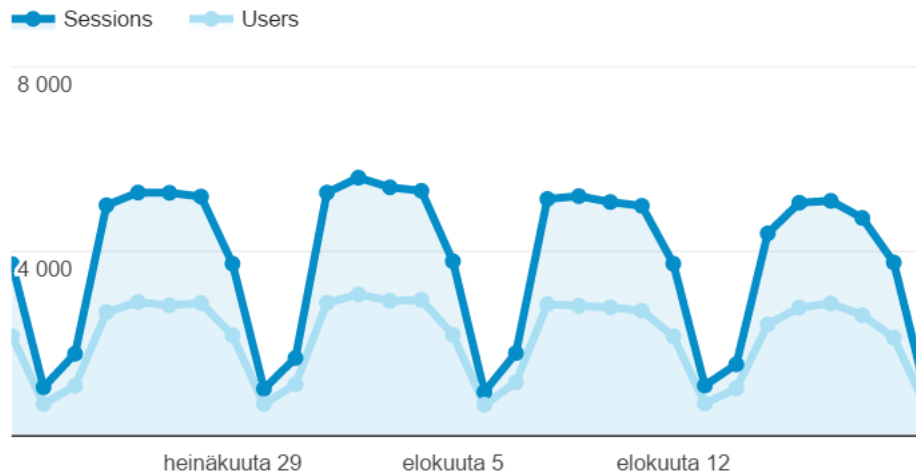
Embed API on JavaScript-kirjasto, jonka avulla voi luoda työpöytäkymiiä tehokkaasti verkkosivuille. Embed API:n avulla voi

- saada yhdellä klikkauksella Google Analytics -sisäänkirjautumispainikkeen, jolloin sitä ei tarvitse itse luoda
- tarkastella profiilin valintapainikkeen avulla dataa useasta lähteestä samaan aikaan

- luoda integraation Google-tilin kanssa, jolloin visuaalisen sisällön luonti helpottuu
- suorittaa omia kyselyitä täysin oikeuksin ja käyttää kolmannen osapuolen työkaluja, kuten mukautettuja kaavioita
- rakentaa omia komponentteja, joita voi uudelleen käyttää ja jakaa muiden kanssa. [19.]

Google Analytics Embed API käyttää Core Reporting API:a hakiessaan dataa Google Analyticsistä. Embed API:iin voi tehdä kyselyitä käyttäen samoja kyselyrajoitteita kuin Core Reporting API:ssa. Embed API:n avulla voi myös sallia palvelinpuolisen auktorisoinnin, jolloin käyttäjän ei tarvitse kirjautua sisään omalla tunnuksellaan.

Embed API:n avulla on esimerkiksi helppo toteuttaa graafinen diagrammi: esimerkki kaaviosta on esitetty kuvassa 7. Kaaviossa on esitetty kuukauden takainen sessio- ja kävijädata Google Analytics Demos & Tools -sivustolla [20].



Kuva 7. Embed API:n avulla toteutettu diagrammi [20].

Multi-Channel Funnels Reporting API mahdollistaa kyselyt todennetulle käyttäjälle. Todennettu käyttäjä on käyttäjä, joka on kirjautunut sisään omalla Google-tilillään. Ohjelmointirajapinnan avulla on mahdollista suorittaa kyselyitä Multi-Channel Funnels -dataan ja -raportteihin. Multi-Channel Funnel -raportit kertovat, mistä lähteistä sivustolle

päätyneet käyttäjät ovat tulleet ja mitä käyttäjät ovat tehneet sivustolla ennen ostopäätöstä tai muuta haluttua lopputoimintoa. Multi-Channels näyttää, kuinka kanavat toimivat yhteen ja lisäävät myyntiä sekä tuottavat konversiota. Multi-Channelsin avulla voidaan tutkia, mitkä kanavat ja avainsanat johtavat myyntiin ja millainen on käyttäjän matka ostopäätöksiin tai haluttuun lopputoimintoon. Multi-Channel Funnels Reporting API:n avulla voi

- rakentaa räätälöityjä raportteja Multi-Channels-datasta
- yhdistää ja esittää Multi-Channel Funnels -dataa yritysdatan kanssa
- hyödyntää dataa ja luoda visuaalisia esityksiä. [21.]

Real Time Reporting API mahdollistaa reaaliaikaisen datan hakemisen Google Analyticsistä kirjautuneille käyttäjille. Real Time Reporting API:n avulla voi

- näyttää aktiiviset käyttäjät sivustolla
- näyttää suosituimmat sivustot reaaliaikaisesti
- luoda reaaliaikaisen työpöydän.

Metadata API palauttaa sarakkeiden (*columns*) listan ja attribuutit eli ulottuvuus ja metriset (*dimensions and metrics*) tiedot. Metadata API:a voidaan käyttää esimerkiksi haluttaessa seuraavia tietoja:

- vanhentuneet sarakkeet
- sarakkeiden nimet
- mallinnetut sarakkeet (*templated*)
- lasketut sarakkeet
- sarakkeet ja segmentit

- API-versiot
- eTagit.

"Vanhentuneet sarakkeet" palauttaa "DEPRECATED", jos sarake on vanhentunut. Tätä voidaan hyödyntää testattaessa, ovatko kentät vanhentuneet tai muuttaneet nimeä Google Analytics ohjelmointirajapinnoissa. "Lasketut sarakkeet" palauttaa arvon riippuen siitä, onko sarake laskettu. Esimerkiksi sarake "ga:percentNewSessions" on laskettu kaavalla "ga:newSession / ga:sessions". "Sarakkeet ja segmentit" tarkoittavat sarakkeen soveltuvuutta esimerkiksi segmentiksi, joka voidaan testata "allowedInSegments" -attribuutin avulla. "ETagit" sisällytetään jokaiseen Metadata API:n vastaukseen. Tätä tunnistetta voidaan käyttää datan säilömiseen välimuistissa tai päivittämään Metadata API:n vastauksia.

Management API mahdollistaa pääsyn Google Analyticsin asetuksiin. Se mahdollistaa suurten kokonaisuuksien hallinnan ja esimerkiksi pääsyn Analyticsin käyttäjätileihin. Google Management API:a voi käyttää

- listaamaan kaikki tili-, ominaisuus- ja näkymätiedot käyttäjälle
- hallitsemaan ominaisuuksia, näkymiä ja tavoitteita
- hallitsemaan käyttäjätasoa ja oikeuksia
- palauttamaan näkymä-ID, jota voidaan käyttää Core Reporting API:ssa
- määrittämään mitkä tavoitteet ovat aktiivisia, ja muokkaamaan näiden asetuksia
- hallitsemaan linkkejä Analyticsin ominaisuuksien ja AdWordsin käyttäjien välillä
- hallitsemaan Remarketing-yleisöä.

Provisioning API:a voidaan käyttää uusien Google Analytics -käyttäjien luontiin ja mahdollistamaan Google Analytics -asiakkaille. Provisioning API on tarkoitettu lähinnä suurille Google-partnereille ja sen käyttö vaatii kutsun lähettämisen ja hyväksynnän ja erityisten ohjelmointirajapintatermien hyväksymisen.

4.2 Ohjelmointirajapintojen rajoitteet

Kaikkien Google Analytics API:en rajoitteet eroavat hieman toisistaan eri käyttötarkoitusten vuoksi. Google on asettanut niin sanotut ”yleiset rajoitteet”, jotka ovat yhteisiä Management API:lle, Core Reporting API:lle, Multi-Channel Funnels API:lle, Metadata API:lle ja Real Time Reporting API:lle. Nämä rajoitteet ovat

- 50 000 kyselyä päivässä projektia kohden
- 10 kyselyä sekunnissa IP:tä kohden. [22.]

Lisäksi yhteisenä rajoituksena ohjelmointirajapinnoille Core Reporting API, Realtime API ja Multi-Channel Funnels API on

- 10 000 kyselyä päivässä näkymää (profiilia) kohden
- 10 samanaikaista kyselyä näkymää (profiilia) kohden.

Profiili tarkoittaa Google Analyticsissä alinta tasoa Analytics-näkymässä. Eri profiileja voidaan luoda eri näkymille. Jos esimerkiksi halutaan joidenkin käyttäjien näkevän vain tiettyjen URL-osoitteiden analytiikkatiedot, tämä voidaan toteuttaa eri profiilien avulla. Projekti on profiilin ylempi taso. Projekti on yksi tili, jolla on yksi seurantatunnus ja jonka alle voi kuulua useita profiileja.

Analytics Reporting API:lle on olemassa erityiset rajoitukset:

- 50 000 kyselyä päivässä projektia kohden
- 2 000 kyselyä 100 sekunnissa projektia kohden
- 1 000 kyselyä 100 sekunnissa käyttäjää ja projektia kohden.

Jos projektirajoitukset ylittyvät Analytics API palauttaa virheviestin statuksella 403. Kyselyille on mahdollista pyytää rajoitusten nostoa. Tämä on tapauskohtausta, ja rajoitusten nostoista päättää Google.

Ohjelmointirajapintojen palauttamiin virheviesteihin pätevät seuraavat rajoitteet:

- 10 virheviestiä tunnissa profiilia ja projektia kohden
- 50 virheviestiä päivässä profiilia ja projektia kohden.

Yleisten rajoitusten mukaan, jos sivulatauksia olisi päivittäin noin 700, voisi kyselyitä tehdä

$$\frac{50000}{700} \approx 71,42$$

Tämä tarkoittaisi, että jokainen käyttäjä voisi päivittää sivun 71 kertaa tai navigoida eri sivuilla 71 kertaa päivässä, jos yhdellä sivulla lähetettäisiin yksi kysely. Jos sivustolla suoritetaan useampia kyselyitä, on kyselyt myös rajoitettava 10 kyselyyn sekunnissa.

5 Palvelun taustatutkimus ja suunnittelu

Insinööriyössä tarkoituksena oli tuottaa rekrytoiville yrityksille palvelu, jossa ne pystyvät jättämään uusia työpaikkailmoituksia ja näkemään työpaikkailmoituksiinsa tai muihin yritykseen liittyviin sivustoihin liittyviä tilastoja. Työpöytä näkymäpalvelu antaa yrityksille selkeän käyttöliittymän, joka auttaa heitä seuraamaan ja analysoimaan työpaikkailmoituskampanjoiden etenemistä ja lukijakuntaa.

Käytän raportissa ohjelmointirajapinnasta lyhennettä API yksinkertaisuuden vuoksi. Suunnitteluvaiheessa laadittiin projektisuunnitelma, jonka mukaan projektissa edettiin. Ratkaistaviksi kysymyksiksi projektin suunnittelussa osoittautuivat seuraavat:

- Ovatko API-kyselyiden rajoitteet liian tiukat?
- Tarvitseeko dataa tallentaa omaan kantaan vai riittääkö reaaliaikainen haku API:sta?
- Jos dataa tallennetaan kantaan, millaista dataa tallennetaan?
- Onko suorituskyky ja nopeus riittävä, jos dataa haetaan suoraan API:sta?

- Kuinka raskaista kyselyistä API pystyy suoriutumaan?

5.1 Tarvittava data ja ohjelmointirajapinnan valinta

Sovelluksessa oli tarkoitus toteuttaa käyttäjälle työpöytä, jossa käyttäjän on mahdollista nähdä yrityksensä tilastotietoja (*statistics*). Yritys voi jättää palveluun työpaikkailmoituksia ja sovelluksen avulla nähdä näiden työpaikkojen tilastotietoja. Sivustolla tulisi näyttää eri ilmoitusten tietoja kuten lukukerrat, ilmoituksessa vierailleiden käyttäjien lähteet ja käytetyt laitteet. Sovellus tarjoaisi myös kokonaiskuvan kuukausitasolla käyttäjän ilmoituksista ja niiden tilastotietoja.

Dataa olisi tarkoitus esittää graafisina elementteinä, kuten visualisina kaaviona, mutta myös taulukoina. Sovelluksen tulisi olla käyttäjäystävällinen ja suorituskyvyltään luotettava. Toteutuksessa tuli ottaa huomioon myös API-rajoitteet.

API:n valintaan vaikuttivat API:n ominaisuudet ja API:sta saatava data. Kriteerinä oli myös datan näyttäminen ilman käyttäjän todentamista Google-tilin avulla. Lopulliseksi valinnaksi tuli Embed API, koska API:sta saatava data vastaa haluttua dataa. Embed API:lla luotavat kaaviot ja palvelinpuolinen käyttäjätodennus vaikuttivat myös Embed API:n käyttöönottoon ja hyödyntämiseen. API:ien dokumentaatiot ovat selkeät, joten ne oli helppo ottaa käyttöön. Embed API käyttää perustanaan Core Reporting API:a, joten osassa tätä työtä vertailtiin myös tuloksia, jotka on saatu Core Reporting API:sta.

5.2 Datan tarkkuus

Datan tarkkuus on myös tärkeä osa toteutusta, koska käyttäjälle halutaan näyttää todennukaista dataa. Markkinointi perustuu pitkälti sivulatauksiin ja lukukertoihin. Kampanjoita myydään lukukertojen perusteella, joten datan tulisi olla todennukaista.

Google Analytics -data ei välttämättä ole niin tarkkaa, kuin sen luulisi olevan. Mitä suuremmiksi sivuston vierailijamäärät kasvavat, sitä epätarkemmaksi data muuttuu. Tämä johtuu siitä, että isojen datamäärien käsittely vie enemmän aikaa, joten Google käyttää pienempää dataotantaa (*sampled data*). Pienempien kävijämäärien sivustojen raportit ovat tarkempia, koska tällöin kaikki kävijädata otetaan mukaan raporteihin. Suurin osa

raporteista ei sisällä dataotantaa vaan on laadittu koko dataosuudesta. Dataotantaa käytetään tilastollisissa analyyseissä, koska pienemmän alaryhmän analysointi antaa lähes samanlaisen tuloksen kuin koko datamäärän analysointi. Pienemmän datamäärän analysointi on myös tehokkaampaa ja tuloksien kyselyt nopeampia. Google Analyticsissä dataotantaa voidaan käyttää joko datan keräysvaiheessa tai raporteissa.

Pääsyy pienemmän dataotannan käyttöön ovat yli 500 000 sessiota sisältävät kyselyt [22]. Jos dataotantaa on käytetty, Google Analytics antaa kuvassa 8 näkyvän huomautuksen raportin yhteydessä.

This report is based on 235,951 sessions (35.71% of sessions). [Learn more](#)

Kuva 8. Google Analyticsin antama huomautus, jos näytedataa on käytetty kyselyssä.

Dataotannan osuutta voi säätää Google Analyticsissä täsmällisemmäksi. Testattaessa esimerkiksi uusien ja palaavien käyttäjien dataa, tarkin mahdollinen dataotannan osuus oli 69,21 %. Jos dataotannan asetti epätarkaksi, sama luku oli 0,14 %. Esimerkki tuloksien eroavaisuuksista on esitettyinä taulukossa 2.

Taulukko 2. Google Analyticsin näytedatan eroavaisuudet.

	Korkea näytedata	50 %	Matala näytedata
Sessiot	631 792	659 864	660 810
Sessiot / Kaikki	95,63 %	99,88 %	100,03 %

Esimerkkitapauksessa eroavaisuutta esiintyi suuren ja pienen dataotannan välillä noin 30 000 sessiota. Pienen dataotannan tapauksessa tuloksen virheellisyyden huomaa myös sessioiden prosentuaalisesta osuudesta, joka väittää sessioita olevan 0,03 % enemmän kuin sessioita yhteensä, mikä käytännössä on mahdotonta. Suuren dataotannan testit osoittautuivat lähes aina ajallisesti noin kaksinkertaisiksi verrattuna hakuun pienellä dataotannalla. Mainitun esimerkin tuloksia testattaessa matalan näytedatan hakuun kului noin 3,52 sekuntia ja korkean näytedatan hakuun noin 6,83 sekuntia.

Core Reporting API:ssa voidaan määritellä myös dataotannan taso (*sampling level*). Parametrille "samplingLevel" voidaan määritellä jokin arvoista

- DEFAULT
- FASTER
- HIGHER_PRECISION.

Oletuksena Core Reporting API käyttää arvoa "DEFAULT". Arvot toimivat samoin kuin Google Analyticsin käyttöliittymässä. "FASTER" palauttaa vastauksen nopeammin, mutta käyttäen pienempää dataotantaa. "HIGHER_PRECISION" palauttaa tarkemman tuloksen käyttäen suurempaa dataotantaa, mutta suoriutuu kyselystä hitaammin. Jos Core Reporting API:n vastaus sisältää vain pienemmän otannan dataa, parametri "containsSampledData" palauttaa "true". Tällöin otantaprosentti voidaan laskea kaavalla

$$\frac{sampleSize}{sampleSpace} * 100 \%$$

Taulukossa 3 ovat esitettynä esimerkkitapaukset eri dataotannoista Google Query Explorerilla [24] ja Google Analytics -sovelluksessa testattuna. Esimerkkidatana haettiin uusien ja palaavien käyttäjien sessiot ja ikäjakauma. Taulukossa on verrattu eri datalähteillä matalan, keskitason ja korkean dataotannan sekä Core Reporting API:n ja Google Analyticsin palauttaman datan eroja.

Taulukko 3. Matalan, korkean ja keskitason dataotannan vertailu eri näytteillä Google Analyticsin ja Core Reporting API:n välillä.

	Matala dataotanta	Keskidataotanta	Korkea dataotanta
Sessiot API (data 1)	118 250	86 801	86 801
Google Analytics (data 1)	122 821	88 719	89 168
Sessiot API (data 2)	857 997	1 073 430	1 073 430
Google Analytics (data 2)	857 997	1 093 662	1 093 587
Sessiot API (data 3)	1 941 218	2 487 055	2 487 055
Google Analytics (data 3)	1 941 218	2 461 351	2 486 621
Sessiot API (data 4)	188 061	198 297	198 297
Google Analytics (data 4)	193 763	199 856	203 076

Matalalla dataotannalla tulokset eroavat huomattavasti keskidataotannan ja korkean dataotannan tuloksista. API:lla tarkasteltuna keskidataotannan ja korkean dataotannan välillä eroa ei ole ollenkaan. Tämä voi johtua dataotannan tarkkuudesta ja haetun datan joukon pienyydestä, jolloin keskidataotanta käytti jo tarkinta mahdollista dataotanta-arvoa. Taulukossa 4 ovat esitettynä prosentuaaliset erot matalan, keskivälin ja korkean dataotannan välillä. Huomattavaa on, että matalalla otannalla datan eroavaisuus voi olla jopa 35 % korkeaan tai keskiväliotantaan. Huomattava eroavaisuus on myös, kuinka ensimmäisellä testidatalla matala osuus oli korkeampi kuin muut otannat, kun taas seuraavissa dataotannoissa matala dataotanta oli aina pienempi.

Taulukko 4. Prosentuaaliset eroavaisuudet matalan, keskitason ja korkean dataotannan sekä Google Analyticsin ja Core Reporting API:n välillä.

	Matala / keskiotanta	Keskiotanta / korkea	Matala / korkea otanta
Sessiot API (data 1)	136,23 %	100 %	136,23 %
Google Analytics (data 1)	138,43 %	99,50 %	137,74 %
Session API (data 2)	79,93 %	100 %	79,93 %
Google Analytics (data 2)	78,45 %	100,01 %	78,46 %
Sessiot API (data 3)	78,05 %	100 %	78,05 %
Google Analytics (data 3)	78,87 %	98,98 %	78,07 %
Session API (data 4)	94,84 %	100 %	94,84 %
Google Analytics (data 4)	96,95 %	98,41 %	95,41 %

Testeissä matalalla dataosuudella tehdyt testit osoittautuivat ajallisesti noin puolta nopeammiksi.

Haluttaessa tarkempaa dataa vaihtoehtona on Google Analytics 360 -tili, jonka avulla esimerkiksi dataotannan raja alkaa vasta 100 miljoonasta sessiosta. Google Analytics 360 on osa Googlen maaliskuussa 2016 julkistamaa Suite 360 -sarjaa, johon kuuluu kuusi eri tuotetta:

- Analytics 360
- Tag Manager 360

- Attribution 360
- Optimize 360 (beta)
- Audience Center (beta)
- Data Studio 360 (beta).

Google Analytics 360:n hintapyyntöä tulee kaikissa tapauksissa pyytää yksilökohtaisesti Googlelta, mutta jotkut lähteet sanovat hinnan alkavan noin 150 000 dollarista vuodessa [25].

5.3 Datan omistajuus ja oikeudet

Dataa tallennettaessa kysymykseksi nousee, kuka datan omistaa. Voiko Google Analyticsista saatavaa dataa käyttää mielivaltaisesti? Google API:en käyttöehdoissa pykälässä 5 d on seuraava maininta datan siirrettävyydestä:

Google supports data portability. For as long as you use or store any user data that you obtained through the APIs, you agree to enable your users to export their equivalent data to other services or applications of their choice in a way that's substantially as fast and easy as exporting such data from Google products and services, subject to applicable laws, and you agree that you will not make that data available to third parties who do not also abide by this obligation. [26.]

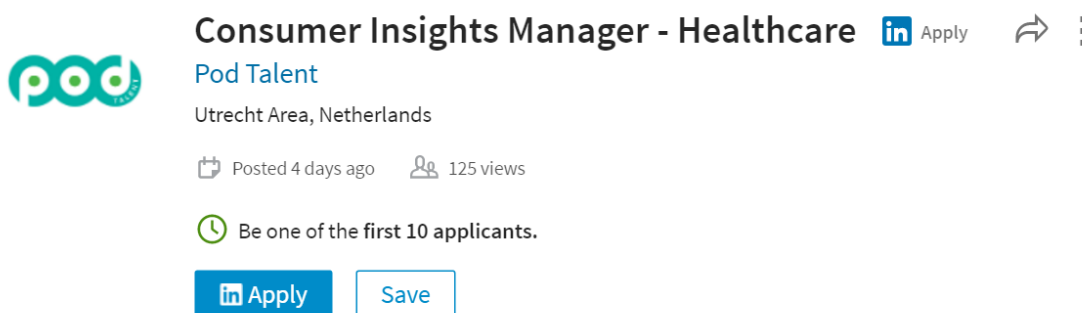
Google tukee datan siirrettävyyttä, jolloin Google API:sta saatavan datan tallentaminen ja myöhemmin käyttäminen on hyväksyttyä.

Google Analyticsin tilin ylläpitäjä myös omistaa tiliin liittyvän Google Analytics -datan. Tiliin liitetyt käyttäjät voivat viedä dataa (*export*) esimerkiksi Google Analytics Reporting API:n avulla. Vietyä dataa voidaan käyttää riippumattomasti ilman Google Analyticsiä tai muita sovelluksia tai palveluita, jotka ovat liittyneitä Googleen. [27.]

5.4 Vastaavat palvelut

Markkinoilla ei ole tullut vastaan täysin vastaavaa palvelua. Markkinoilta löytyy kuitenkin joitakin palveluja, jotka tarjoavat jätetystä työpaikkailmoituksesta syvempää tietoa (*insights*).

Esimerkiksi verkostoitumispalvelu LinkedIn tarjoaa tarkempaa tietoa jätetystä työpaikkailmoituksesta ja mahdollisuuden nähdä, miten ilmoitus suoriutuu verrattuna muihin ilmoituksiin tai millaisia hakijoita ja katselijoita se on saanut [28]. LinkedIn tarjoaa tietoa ilmoituksista myös peruskäyttäjille. Esimerkiksi kuvassa 9, joka on ote LinkedIn-palvelusta olleesta työpaikkailmoituksesta, ovat nähtävissä työpaikkailmoituksen katselukerrat (*views*).



Kuva 9. Osa työpaikkailmoituksesta LinkedIn-sivustolla [29].

LinkedIn tarjoaa myös lisäpalvelua, mutta se ei paranna verkkoanalytiikkaan liittyviä palveluja työnhakijalle. Premium-palvelun avulla käyttäjä voi nähdä esimerkiksi millaisia hakijoita työpaikkaa on jo hakenut.

5.5 Riskit

Jokaiseen projektiin liittyy omat riskinsä. Teknisellä puolella API-päivitykset tai palvelun lopetus tulee huomioida. Core Reporting API julkaistiin toukokuussa 2005. Tämän jälkeen uusia, pienempiä versioita on julkaistu noin 1–3 kuukauden välein. Viimeisin isompi julkaisu, versio 4.0, ilmestyi huhtikuussa 2016. Sitä aikaisempi julkaisu, versio 3.0, julkaistiin kaksi vuotta aikaisemmin, huhtikuussa vuonna 2014. Embed API julkaistiin toukokuussa 2014, eikä siihen ole tämän jälkeen tullut päivityksiä. [30.]

Voidaan olettaa, että Core Reporting API tulee päivittymään 1–3 kuukauden välein. Nämä päivitykset tulevat kuitenkin olemaan pieniä, eivätkä ne vaikuta suuresti API:n toimintakykyyn tai aikaisempien ominaisuuksien toimintaan. Yleisimmin päivityksien yhteydessä julkaistaan uusia ominaisuuksia tai korjataan mahdollista ohjelmointivirheitä. Uusista päivityksistä olisi kuitenkin hyvä olla tietoinen, ja niitä olisi hyvä seurata noin 1–2 kuukauden välein. Useimmiten pienet päivitykset eivät vaadi oman sovelluksen päivittämistä.

Isot päivitykset tapahtuvat aikaisempien päivitysten perusteella noin kahden vuoden välein. Näihin päivityksiin on hyvä reagoida viimeistään 6 kuukauden kuluttua julkaisusta, jos päivitys vaikuttaa omaan sovellukseen. Google on asettanut kuuden kuukauden minimitukiajan poistuville ominaisuuksille esimerkiksi siirryttäessä versiosta 3.0 versioon 4.0 [31]. Vaikka päivitykset eivät koskisi omaa sovellusta, on niihin hyvä kuitenkin reagoida ja päivittää API ajan tasalle. Tällöin seuraaviin päivityksiin siirtyminen on helpompaa kuin usean version ylitse siirtyminen. Google ei kuitenkaan lopeta sovellusten tukea nopeasti, joten käyttäjillä on aikaa reagoida muutoksiin.

Googlea voidaan myös yleisesti pitää luotettavana palveluntarjoajana, joten Google Analyticsin ja siihen liittyvien API:en voidaan olettaa pysyvän toiminnassa ja niihin liittyvistä muutoksista raportoitavan. Palveluilla on myös useita käyttäjiä, ja Googlella yli 100 erilaista API:a, jotka ovat nähtävissä Google Developers Consolesta (URL: <https://console.developers.google.com>).

Jotta API:n päivityksiin pystytään reagoimaan, on hyvä päättää, kuka on vastuussa päivitysten tarkastuksesta ja mahdollisista muutoksista sovellukseen. API:n toteutus on hyvä dokumentoida, jotta mahdollisesti eri henkilöt pystyvät helposti hoitamaan ongelmia ja päivityksiä.

5.6 Suunnittelun tulos

Insinöörityössä päädyttiin suunnittelun tuloksena käyttämään sekä API:a että tallentamaan osa datasta omaan tietokantaan. Tähän vaikuttivat sekä API-rajoitteet että API-kyselyiden hitaus. Aluksi tarkoituksena oli tallentaa dataa työpaikkailmoitusten yhteyteen kuukausitasolla. Lopulta päädyttiin ratkaisuun, jossa lukukerrat tallennetaan päivä-

tasolla. Tällöin pystytään näyttämään tilastodiagrammi, jossa yrityksen kaikkien ilmoitusten lukutilastot näytetään päivätasolla. Samalla päivätasolla tallennetaan myös hakijaklikkaukset, joita tarvitaan kuukausiraporteissa.

Data päädyttiin tallentamaan työpaikkailmoituksiin, jotka ovat sulkeutuneet. Tällöin avoinna olevien ilmoitusten data tulee olemaan reaaliaikaista. Myös yksittäisen ilmoituksen tilastoja tarkasteltaessa data haetaan reaaliaikaisesti. Data työpaikkailmoituksiin tallennetaan päivittäin.

Suorituskyky osoittautui tarpeeksi nopeaksi, vaikka data olisi haettu suoraan API:sta reaaliaikaisesti. Kyselyn kesto vaihteli välillä 2–5 sekuntia. Tämä toimi tarpeeksi hyvin, koska kysely suoritettiin vasta sivuston jo latauduttua, jolloin lataus ei pysäyttänyt koko sivuston latautumista. Suorituskykyä kuitenkin testattiin vielä toteutusvaiheessa.

API-kyselyiden tuloksissa tulee myös huomioida, että työpaikkailmoitusten URL-rakenne voi olla kahta eri muotoa:

- /<id>
- /xxx-xxx-<id>.

6 Työpöytänäköpalvelu

Toteutuksissa käytetään Google Analyticsin Embed API:a. Sen avulla palvelinpuolinen käyttäjän todennus onnistui helposti, jolloin dataa voidaan näyttää ilman käyttäjän kirjautumista. Palvelinpuoliseen käyttäjän todennukseen on kolme vaihetta:

1. palvelutilin luonti ja JSON-avaimen lataus
2. palvelutilin asetus käyttäjäksi Google Analyticsissä
3. yhteysavaimen (*access token*) pyytäminen JSON-avaimen avulla [1].

Palvelutilin luonti, JSON-avaimen lataus ja palvelutilin käyttäjäksi asetus onnistui helposti Googlen luomilla ohjeilla [20] Google Developers Consolin (URL: <https://console.developers.google.com>) kautta. Yhteysavain pyydetään Pythonilla kuvan 10 mukaisesti.

```

from oauth2client.service_account import ServiceAccountCredentials

# The scope for the OAuth2 request.
SCOPE = 'https://www.googleapis.com/auth/analytics.readonly'

# The location of the key file with the key data.
KEY_FILEPATH = 'path/to/json-key.json'

# Defines a method to get an access token from the ServiceAccount object.
def get_access_token():
    return ServiceAccountCredentials.from_json_keyfile_name(
        KEY_FILEPATH, SCOPE).get_access_token().access_token

```

Kuva 10. Yhteysavaimen pyyntö JSON-avaimen avulla [20].

Yhteysavaimen hankinnan jälkeen sitä voidaan käyttää kyselyissä Embed API:lle. Jokaisessa näkymässä, jossa kyselyitä halutaan luoda, tulee Embed API -kirjasto ladata ennen API:n kutsumista. Embed API:n latauskoodi on kuvattuna kuvassa 11.

```

<script>
(function(w,d,s,g,js,fs){
  g=w.gapi||(w.gapi={});g.analytics={q:[],ready:function(f){this.q.push(f)}};
  js=d.createElement(s);fs=d.getElementsByTagName(s)[0];
  js.src='https://apis.google.com/js/platform.js';
  fs.parentNode.insertBefore(js,fs);js.onload=function(){g.load('analytics')}};
})(window,document,'script');
</script>

```

Kuva 11. Embed API -kirjaston latauskoodi.

Lähetettäessä kyselyitä API:lle on ensin varmistettava, että kirjasto on latautunut koodilla

```
gapi.analytics.ready(function() {});
```

Tämän jälkeen käyttäjä todennetaan yhteysavaimella kuvan 12 mukaisesti.

```

gapi.analytics.auth.authorize({
  'serverAuth': {
    'access_token': '{{ ACCESS_TOKEN_FROM_SERVICE_ACCOUNT }}'
  }
});

```

Kuva 12. Käyttäjän todennus yhteysavaimella.

Kun käyttäjä on todennettu, voidaan API:in luoda eri kyselyjä. Toteutus koostuu kokonaisuudessaan kolmesta eri näkymästä, joilla on tarkoitus esittää Google Analyticsiin perustuvaa dataa. Näkymät ovat

- ilmoitussivu
- yksittäisen ilmoituksen sivu
- tilastosivu.

Kaikki datakyselyt Embed API:lle tehdään samalla kaavalla, joka on esitettyä kuvassa 13.

```

var example = new gapi.analytics.report.Data({
  query: {
    'ids': 'ga:98765432',
    'start-date': '2016-05-01',
    'end-date': '2016-06-01',
    'metrics': 'ga:pageviews',
    'dimensions': 'ga:previousPagePath',
    'filters': 'ga:landingPagePath=~tyopaikat/tyo.*?' + id
  }
});
example.on('success', function(response) {

})
example.execute();

```

Kuva 13. Esimerkkikysely Embed API:lle.

Parametrit ja arvot määritellään "query"-parametrille, ja vastaus käsitellään "success"-funktiolla sekä suoritetaan "execute"-funktiolla.

6.1 Palvelun sivunäkymät

Ilmoitussivu

Ilmoitussivulla oli tarkoitus esittää yrityksen työpaikkailmoitukset ja niiden näyttökerrat. Google Analyticsista haettava data on kokonaisnäyttökerrat jokaiselle ilmoitukselle ilmoituksen aukioloaikana. Näkymässä päädyttiin ratkaisuun, jossa näyttökerrat haetaan vain avoinna oleville ilmoituksille suoraan API:sta reaaliaikaisesti. Sulkeutuneille ilmoituksille näyttökerrat on tallennettu valmiiksi tietokantaan.

Näkymässä päädyttiin käyttämään myös Data Tables -kirjastoa [32] taulukon lajitteluun ja hallitsemiseen. Tämän vuoksi data päädyttiin hakemaan aina näkyvillä oleville työpaikkailmoituksille. Ensin näytetään kymmenen ensimmäistä ilmoitusta, joille haetaan näyttökerrat heti, kun sivu on latautunut. Koska tulokset haetaan vain kymmenelle ensimmäiselle, on taulukon lajittelu näyttökertojen perusteella mahdotonta. Tämä ratkaisu kuitenkin hyväksyttiin, koska jos näyttökerrat olisi haettu kaikille ilmoituksille, suoritusnopeus olisi laskenut huomattavasti. Huomioitavaa oli myös, että joillakin yrityksillä on tuhansia ilmoituksia ja satoja avoinna olevia työpaikkailmoituksia. Jos näyttökerrat haettaisiin aina kaikille avoinna oleville ilmoituksille, voi API suorittaa yli 100 kyselyä yhdellä sivun päivityksellä, jolloin API-rajat tulisivat helposti vastaan.

Suoritettu kysely yhdelle työpaikkailmoitukselle koostuu seuraavista parametreista, jotka ovat esitetynä taulukossa 5. Suodattimena (*filter*) toimiva "ga:pagePath" on säännöllinen lauseke (*regular expression*), jossa arvon "ga:pagePath" on sisällettävä merkkijono "työpaikat/työ". Tämän jälkeisillä merkeillä ei ole väliä, mikä kerrotaan merkkijohdistelmällä ".*?". Tämän jälkeen arvon on vielä sisällettävä "id", joka kuvaa jokaiselle työpaikkailmoitukselle yksilöllistä numeroyhdistelmää.

Taulukko 5. Embed API:lle lähetettävät parametrit haettaessa näyttökertoja.

Parametri	Arvo
metrics	ga:pageviews
dimensions	ga:pagePath
filter	ga:pagePath=~tyopaikat/tyo.*? <id>

Kysely palauttaa kaikki mahdolliset URL-osoitteet, jotka osuivat hakuun, sekä niiden näyttökerrat. Esimerkkipalautus, joka on saatu Query Explorerilla [24], on esitettyä kuvassa 14.

Page	Pageviews
duunitori.fi/tyopaikat/tyo/2600007	93
duunitori.fi/tyopaikat/tyo/inlook-tyonjohtaja-helsinki-sdsuu-2600007	131

Kuva 14. Esimerkkipalautus työpaikkailmoituksesta.

Säännöllisellä lausekkeella ratkaistiin ongelma, jossa kahden erilaisen URL-rakenteen tulokset tuli molemmat sisällyttää mukaan kyselyyn. Embed API:n palauttamista tuloksista kokonaislukukerrat voidaan hakea kuvan 15 mukaisella koodilla.

```
totalPageViews = result.response.totalsForAllResults['ga:pageviews'];
```

Kuva 15. Kokonaislukukertojen parsiminen tuloksista.

Näyttökerrat latautuvat API:n kautta reaaliaikaisesti näyttökerrat-sarakkeeseen aina sivupäivityksen yhteydessä. Jos taulukkoa päivitetään ja rivejä, joille näyttökertoja ei ole vielä haettu, ilmaantuu lisää, haetaan myös niille näyttökerrat.

Embed API:n datarajoitus ”10 kyselyä sekunnissa per IP” osoittautui rajoitteeksi, jonka vuoksi tulosten latausta oli myös hidastettava. Alkunäkymässä, jossa taulukon sarakkeita on 10, ei ongelmaa ole, mutta jos taulukon päivittää näyttämään esimerkiksi 100 riviä kerralla, on API:a kutsuttava myös 100 kertaa. Ratkaisuna kyselyjä hidastettiin, jolloin API:a kutsutaan vain 100 millisekunnin välein, jolloin maksimi kyselymäärä sekunnissa on 10.

Myöhemmin ratkaisuun tehtiin vielä muutos, koska erillisten kyselyiden lähettäminen saattaa aiheuttaa liian monia kyselyitä. Muutoksessa luvut päätettiin hakea yhdellä kyselyllä kaikille näkyvillä oleville riveille. Muutoksena filter-parametrille annetaan useita id-arvoja, jolloin filter-parametri on seuraavan näköinen:

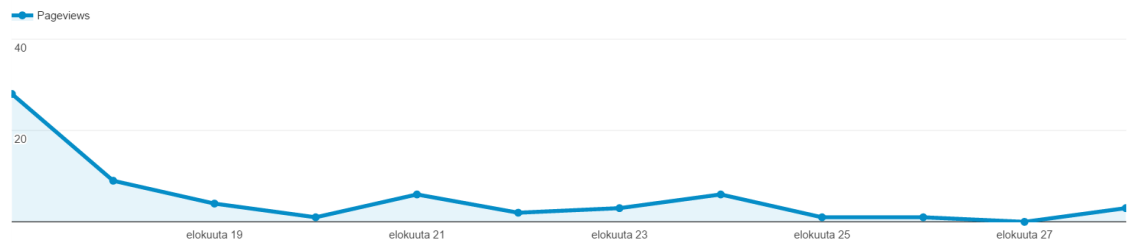
```
ga:pagePath = ~tyopaikat/tyo.*? < id1 >, ga:pagePath = ~tyopaikat/tyo.*? < id2 > ...
```

Kyselylle voidaan antaa vain yksi alku- ja loppupäivämäärä. Aikaisemmin kyselylle annettiin aina kyseisen työpaikkailmoituksen alku- ja loppupäivämäärä. Uudessa kyselyssä

päädettiin ratkaisuun, jossa kaikista kyselyyn osallistuvista riveistä määritetään varhaisin alkupäivämäärä ja vastaavasti myöhäisin loppupäivämäärä.

Yksittäisen ilmoituksen sivu

Yksittäisen ilmoituksen sivulla on tarkoitus esittää ilmoituksen lukukerrat ilmoituksen voimassaoloaikana ja yksityiskohtaisempaa dataa, kuten ilmoitukseen saapuneiden käyttäjien lähteet, käytetyt laitteet ja hakuklikit. Yksittäisen ilmoituksen tapauksessa data päätettiin haettavaksi aina Embed API:n kautta, koska dataa käyttäjien lähteistä tai laitteista ei ollut järkevää tallentaa tietokantaan ja tämä data oli helposti haettavissa API:sta. Data on myös aina reaaliaikaista, jos se haetaan API:sta, jolloin avoimena olevista ilmoituksista on aina saatavilla viimeisin ja ajantasaisin data. Kuvassa 16 on esitettyä Embed API:n kautta piirretty tilastodiagrammi ilmoituksen lukukerroista.



Kuva 16. Yksittäisen ilmoitussivun ulkonäkö.

Tilastodiagrammi on hyvin yksinkertainen toteuttaa Embed API:n avulla, koska se antaa suoran rajapinnan "DataChart" -tilastodiagrammin piirtoon. Samalla API-kyselyllä on mahdollista saada myös kokonaiskatselukerrat. API-kysely on kuvattuna kuvassa 17.

```

var dataChart = new gapi.analytics.googleCharts.DataChart({
  query: {
    'ids': 'ga:██████████',
    'start-date': date_posted,
    'end-date': date_ends,
    'metrics': 'ga:pageviews',
    'filters': 'ga:pagePath=~tyopaikat/tyo.*?' + id,
    'dimensions': 'ga:date',
  },
  chart: {
    'container': 'chart-container',
    'type': 'LINE',
    'options': {
      'width': '100%'
    }
  }
});
dataChart.execute();

```

Kuva 17. Embed API -kysely tilastodiagrammin piirtämiseksi.

Käyttäjien lähtösivut jaottuvat Google Analyticsissa eri tavoin. Jos halutaan tarkastella, onko käyttäjä saapunut sivulle palvelun sisältä vai ulkopuolelta, tulee sivustoa tarkastella landingPage-parametrin avulla ja sivua, jolta käyttäjä on saapunut, previousPath-parametrin avulla. Tämä parametri ei kuitenkaan kerro, mistä ulkopuolisesta sivusta käyttäjä on saapunut. Kun sivustoa tarkastellaan pagePath-parametrin avulla, voidaan ulkopuoliset lähteet nähdä sourceMedium-parametrin avulla. Nämä parametrit eivät kerro sivuston sisäpuolelta tulevia lähteitä, ja tämän vuoksi lähteet on haettava usean eri parametrin avulla ja suoritettava kaksi erillistä kyselyä.

Sivustolla vierailleen käyttäjän laite on helposti jäljitettävissä Google Analyticsin avulla. Tämä voidaan tehdä käyttämällä taulukon 6 osoittamia arvoja.

Taulukko 6. Embed API:lle lähetettävät parametrit haettaessa vierailijoiden käyttämiä laitteita.

Parametri	Arvo
metrics	ga:pageviews
dimensions	ga:deviceCategory
filter	ga:pagePath=~tyopaikat/tyo.*? <id>

Linkkiklikit Google Analyticsista saadaan Tapahtumat-osiosta (*Events*). Embed API -kyselyyn lähetettävät parametrit ovat kuvattuna taulukossa 7.

Taulukko 7. Embed API:lle lähetettävät parametrit haettaessa vierailijoiden klikkauksia eteenpäin.

Parametri	Arvo
metrics	ga:totalEvents
dimensions	ga:eventCategory
filter	ga:pagePath=~tyopaikat/tyo.*? <id>

Kyselystä saatavan tuloksen käsittely eroaa hieman muista tuloksista, koska kysely palauttaa kaikki määritellyt klikkaustapahtumat, joita Google Analyticsissa on määritelty jäljittäväksi. Olin määritellyt jo aiemmin "Click Apply Job" -tapahtuman jäljittämään työpaikkailmoituksista ulospäin meneviä "Hae työpaikkaa" -klikkauksia. Kyselystä saatavasta tuloksesta tarkastetaan, löytyykö parametrilla "Click Apply Job" klikkauksia.

Sivulla eniten hankaluuksia aiheuttivat lähteet, koska Google Analyticsista lähteiden suurpiirteinen tarkastelu on helppoa, mutta jos datan halutaan olevan tarkkaa, täytyy monia tekijöitä ottaa huomioon.

Tilastosivu

Tilastosivulla oli tarkoitus esittää kuukausikohtaista dataa esimerkiksi jätetyistä ilmoituksista, lukukerroista ja hakuklikeistä. Sivustolla päädyttiin käyttämään kannasta haettavaa dataa, koska tarkan datan saaminen Google Analyticsista vaatisi useita kyselyitä, jolloin API-rajoitteet saattaisivat tulla vastaan. Tilastosivu on ainoa sivu, joka ei tee kysyjä Google Analyticsiin. Tilastosivulla oli kuitenkin tarkoitus käyttää ulkoasultaan samantyyppistä datakuvaajaa, jotta käyttökokemus pysyy hyvänä.

Google Embed API:lla pystyy yksinkertaisesti luomaan tilastodiagrammin, joka noudattaa tiettyä ulkoasua. Diagrammi käyttää perustanaan Google Charts -kirjastoa ja Area Chart -tyyppistä diagrammia [33]. Tilastosivun diagrammi toteutettiin saman kirjaston avulla ja konfigurointiin näyttämään samalta kuin Embed API:n luoma diagrammi.

6.2 Virheiden hallinta

API-rajoitteet aiheuttivat eniten ongelmia eri osa-alueilla läpi projektin. Tällöin myös virheviestien käsittelyyn tuli kiinnittää huomiota, koska esimerkiksi rajoitteen ylityessä API palautti virheviestin. Taulukossa 8 on esitettyä projektin kannalta oleelliset virheviestikoodit, virheiden syyt, selitykset ja suositellut jatkotoimenpiteet [34].

Taulukko 8. Embed API:n palauttavat virheviestit.

Koodi	Syy	Selitys	Suosittu jatkotoimenpide
403	dailyLimitExceeded	Käyttäjä on ylittänyt päiväkohtaisen kyselyrajan profiilia tai projektia kohden.	Älä toista ilman korjaustoimenpiteitä.
403	userRateLimitExceeded	Käyttäjän IP-kohtainen rajoite on ylittynyt.	Toista eksponentiaalisesti.
403	rateLimitExceeded	Globaali tai projektin kokonaisrajoitus on ylittynyt.	Toista eksponentiaalisesti.
403	quotaExceeded	10 samanaikaisen kyselyn raja profiilia kohden on ylittynyt.	Toista eksponentiaalisesti.
429	RESOURCE_EXHAUSTED AnalyticsDefaultGroupCLIENT_PROJECT-1d	Projektikohtainen kyselyiden päivärajoitus on ylittynyt.	Älä toista ilman korjaustoimenpiteitä.
429	RESOURCE_EXHAUSTED AnalyticsDefaultGroupCLIENT_PROJECT-100s	Projektikohtainen kyselyiden raja 100:aa sekuntia kohden on ylittynyt.	Toista eksponentiaalisesti.
429	RESOURCE_EXHAUSTED AnalyticsDefaultGroupUSER-100s	Käyttäjäkohtainen kyselyiden raja 100:aa sekuntia kohden on ylittynyt.	Toista eksponentiaalisesti.

Ekspontiaalinen toisto tapahtuu seuraavan kaavan mukaan:

1. Kysely API:lle.
2. API vastaa virheviestillä.
3. Odota 1 sekunti + satunnainen millisekuntimäärä.
4. Toista kysely.

5. API vastaa virheviestillä.
6. Odota 2 sekuntia + satunnainen millisekuntimäärä.
7. Toista kysely.
8. API vastaa virheviestillä.
9. Odota 4 sekuntia + satunnainen millisekuntimäärä.
10. Toista kysely.
11. API vastaa virheviestillä.
12. Odota 8 sekuntia + satunnainen millisekuntimäärä.
13. Toista kysely.
14. API vastaa virheviestillä.
15. Odota 16 sekuntia + satunnainen millisekuntimäärä.
16. Toista kysely.
17. Jos virheviesti vielä toistuu, lopeta ja kirjaa virhe. [34.]

Satunnainen millisekuntimäärä tarkoittaa satunnaista lukua, joka on yhtä suuri tai pienempi kuin 1 000. Sovelluksessa virheviestit käsiteltiin tämän Googlen suositteleman kaavan mukaisesti. Poikkeuksena ovat etusivulle ladattavat kyselyt, joita hidastettiin jo valmiiksi, koska tiedossa oli yli 10 kyselyä sekuntia kohden. Virheviestit tuli kuitenkin ottaa huomioon, koska esimerkiksi usean ihmisen käyttäessä sovellusta yhtä aikaa voivat rajat tulla vastaan odottamattomasti.

7 Testaus ja ongelmat

Suuremmat ongelmat projektissa pystyttiin välttämään. API-rajoitteet aiheuttivat ongelman ladattaessa katselukertoja usealle ilmoitukselle kerralla. Katselukerrat näille ilmoituksille haettiin ensin ilmoituskohtaisesti jokaiselle ilmoitukselle. Tämä aiheutti useita kyselyjä. Ongelma ratkaistiin hakemalla katselukerrat usean URL-osoitteen perusteella ja parsimalla katselukerrat yksittäisille ilmoituksille id:n perusteella URL-osoitteista. Näin suoritetaan vain yksi kysely API:in, jolloin API-rajat eivät ylitä.

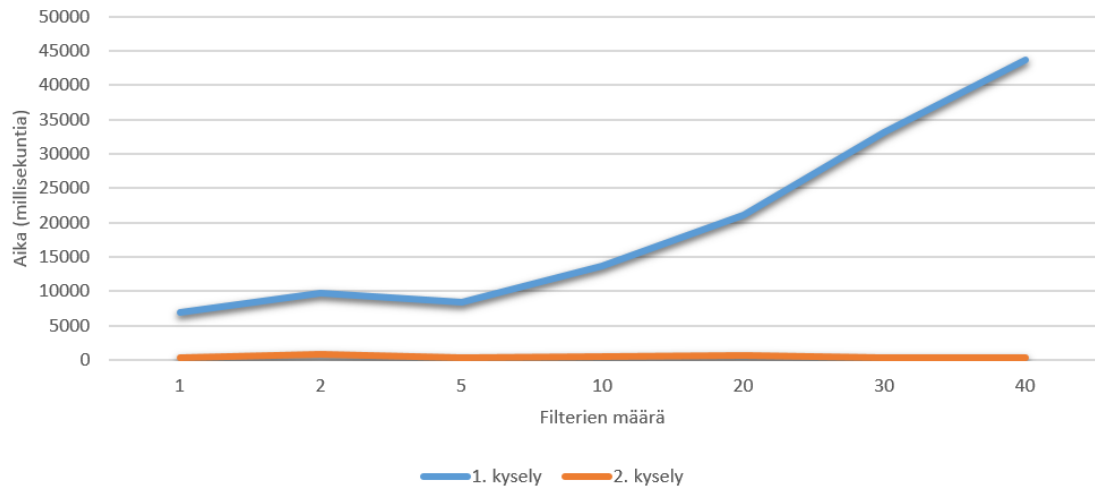
Embed API loi tilastodiagrammin automaattisesti englanninkieliseksi, eikä kyselyparametreilla ollut mahdollisuutta muuttaa kieltä. Google Charteissa kieltä on mahdollisuus muuttaa tilastodiagrammin luontivaiheessa, mutta koska Embed API luo diagrammin automaattisesti omilla arvoillaan, on kieltä hankala muuttaa. Ratkaisuksi päädyttiin kääntämään englanninkieliset sanat JavaScriptin avulla tilastodiagrammin luonnin jälkeen.

Testausta tehtiin läpi projektin, ja sivusto laitettiin sisäiseen testaukseen. Automaatiotestejä projektille ei vielä tehty vaan testejä suoritettiin manuaalisesti. Embed API:a testattiin suurilla datamäärillä ja kyselyillä. Käyttöliittymää testattiin eri selaimilla ja laitteilla niin mobiili- kuin työpöytäversioilla.

Embed API:n suorituskykyä testattiin erikokoisilla kyselyillä. Testit suoritettiin 1 ja 2 vuoden testiajalla ja lähettämällä API:lle eri määrä suodattimia (*filters*). Suodattimien määrä tarkoittaa suodattimeksi asetettuja id-arvoja. Suodatin noudattaa seuraavaa kaavaa:

$$ga:pagePath = \sim tyopaikat/tyo.*? < id >, ga:pagePath = \sim tyopaikat/tyo.*? < id >$$

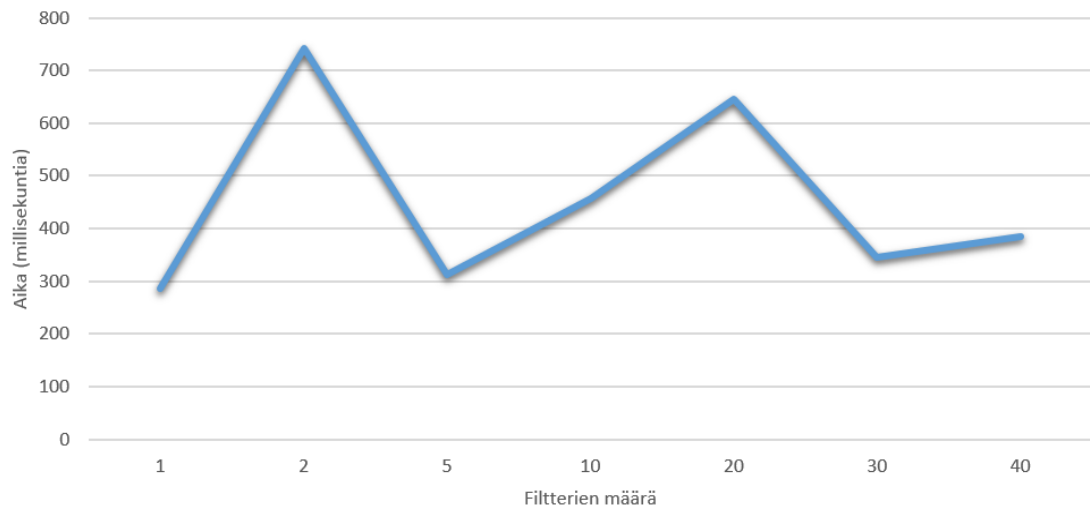
Lauseketta voidaan toistaa useita kertoja eri id-arvoilla. Pilkulla erotettaessa tämä tarkoittaa "OR"-lauseetta, jolloin suodatin luetaan "ga:pagePath sisältää säännöllisen lausekkeen 'tyopaikat/tyo.*?<id1>' tai ga:pagePath sisältää säännöllisen lausekkeen 'tyopaikat/tyo.*?<id2>'". Tuloksista laskettiin keskiarvo. Kaikki tulokset osoittautuivat kuitenkin noudattavan samaa kaavaa. Kuvassa 18 on nähtävissä keskiarvo kahdesta suoritetusta kyselystä. Sininen viiva kuvastaa ensimmäistä kyselyä, joka API:lle lähetettiin. Oranssi viiva kuvastaa kyselyä, joka lähetettiin heti tämän perään, noin 5–10 sekunnin kuluttua.



Kuva 18. Embed API:lle lähetetyn kyselyn kesto millisekunteina eri suodatinmääriä kohden.

Kuten kuvasta 18 on huomattavissa, ensimmäisen kyselyn kohdalla kyselyn kesto aika kasvaa suodattimien määrän kasvaessa. Kuvassa enimmäissuodatinmäärä on 40. Testi tehtiin myös käyttäen 50:tä suodatinta, mutta ensimmäisessä testissä Embed API palautti HTTP-virheen 500 (Internal Server Error). Toinen kysely API:lle onnistui ja vastausaika oli 39 518 millisekuntia. Toisessa testissä molemmat kyselyt onnistuivat. Ensimmäisen kyselyn vastausaika oli 39 072 millisekuntia ja toisen 311 millisekuntia. Ensimmäisen kyselyn HTTP-virhe 500 johtui luultavasti kyselyn aikaväliltä, joka oli ensimmäisessä kyselyssä kaksi vuotta. Pidempien aikavälien kyselyt ovat myös raskaampia Embed API:lle, koska API joutuu hakemaan dataa pidemmältä aikaväliltä.

Huomioon otettavaa oli, että toisen kyselyn suorittaminen samoilla arvoilla heti ensimmäisen kyselyn jälkeen oli huomattavasti nopeampaa. Toinen kysely ei myöskään noudata samaa kaavaa, jossa kyselyn kesto kasvaa suodattimien määrän kasvaessa. Kuvassa 19 on nähtävissä tarkennettu kaavio toisesta kyselystä.



Kuva 19. Embed API:lle lähetettävän toisen kyselyn kesto millisekunteina eri suodatinmääriä kohden.

Toisen kyselyn kesto vaihteli välillä 280–750 millisekuntia. Tulokset eivät olleet riippuvaisia suodattimien määrästä. Toisen kyselyn nopeus johtuu selaimen välimuistiin tallennetuista HTTP-kyselyistä. Selain tallentaa välimuistiin väliaikaisesti tiedostoja, kuten HTML-tiedostoja, CSS-tiedostoja, JavaScript-tiedostoja ja kuvia. Kun sivusto uudelleen ladataan, selain pystyy hakemaan muuttumattomat tiedostot välimuistista, jolloin sivun lataus on nopeampi.

Testaus vaikutti suodattimien määrän käyttöön. Rajoitukseksi päätettiin laittaa 20 suodatinta. Suodatinmäärä haluttiin pitää mahdollisimman isona, jotta API:lle lähetettävä kokonaiskyselymäärä pysyy pienenä. Samalla kuitenkin kyselyn tulisi toimia tarpeeksi tehokkaasti, jottei käyttäjän tarvitse odottaa tuloksia liian kauan.

8 Yhteenveto

Insinööriö antoi laajan kokonaiskuvan verkkoanalytiikasta ja sen käytöstä. Analytiikka sitoo useita osastoja yrityksessä yhteen, ja sen ymmärtäminen on lähes välttämätöntä. Sen avulla voidaan ymmärtää asiakaskuntaa ja toteuttaa parempia palveluja. Analytiikka ei myöskään rajoitu vain yrityksen sisäiseen käyttöön, vaan sitä voidaan hyödyntää myös asiakkaille. Insinööriöprojektin tarkoituksena oli hyödyntää analytiikkaa niin sisäisesti kuin myös käyttäjille.

Työ alkoi tiedon haulla ja käytettävien menetelmien suunnittelusta. Kokonaiskuva projektista hahmoteltiin ja projektisuunnitelma laadittiin. Projektissa pyrittiin huomioimaan tulevat käyttäjät ja heidän käyttökokemuksensa. Koska projekti oli osa isompaa kokonaisuutta, testausta suoritettiin lopussa vasta hieman ja sitä tehdään enemmän koko projektin loputtua.

Projekti sujui kokonaisuudessaan suunnitellusti, mutta hieman aikataulustaan venyen. Tarkkaa aikarajaa ei kuitenkaan ollut asetettu, joten käytännössä venymiselle annettiin tilaa eikä sitä käytännössä edes huomioitu. Projektin oheen ilmaantui myös muita hoidettavia asioita, jolloin venyminen ei johtunut vain projektin itsensä venymisestä. Tekniikkoihin tutustuminen ja päätöksen tekeminen veivät oman aikansa. Google Analyticsin tutustumiseen käytettiin aikaa, koska tiedot perustuivat sen sisältämään dataan. Datan tuli olla oikeellista.

Itse toteutusvaihe sujui nopeasti ja odotetusti. Ainoastaan ohjelmointirajapintarajoitteet vaativat hieman odotettua kauemmin aikaa ja tutustumista rajapintaan paremmin. Kokonaisuutena myös Google Analytics -ohjelmointirajapinnat ovat iso kokonaisuus, ja Analytics on luonut monta erilaista ohjelmointirajapintaa eri tarkoituksiin. Niiden hahmottaminen ja kokonaisuuden näkeminen veivät oman aikansa suunnitteluprosessissa.

Projektien läpivieminen kokonaisuutena on aina oma prosessinsa. Tässä projektissa huomioitavaa oli myös, että välillä projektin joutui keskeyttämään, koska jokin muu ongelma ilmaantui ja se oli korjattava. Käsissä on pidettävä useita lankoja ja uusiin tilanteisiin tai muutoksiin on mukauduttava. Projekti oli kokonaisuutena hyvin nopeatempoinen eikä suunnitteluun ja mietintään ollut hyvin paljoa aikaa ennen toteutusta. Projektissa suunnittelu tehtiin myös jo osana toteutusta ja toteutustapaa saatettiin vaihtaa nopeastikin. Hyvän pohjatyön ansiosta isoja muutoksia ei kuitenkaan ilmaantunut ja pienet oli helppo mukauttaa projektiin.

Verkkoanalytiikkaan tutustuminen antoi uusia näkökulmia ja auttoi ymmärtämään analytiikan todellisen merkityksen. Analytiikka on sidoksissa useisiin eri osastoihin, ja se vaikuttaa useaan osaan yrityksistä. Insinööriyössä tutustuin siihen, kuinka tämä data saadaan esitettyä oikeassa muodossa Google Analyticsistä.

Työssä testattiin Analytics-ohjelmointirajapintojen luotettavuutta ja toiminnollisuuksia. Pääasiassa testaus kohdistui Google Analytics Embed API:in, koska tämä ohjelmointi-

rajapinta valikoitui käytettäväksi sovelluksessa. Embed API osoittautui luotettavaksi ohjelmointirajapinnaksi, joka antaa yksinkertaiset rajapinnat eri toiminnoille. Embed API:a voi hyödyntää niin sisäisessä kuin ulkoisessakin toiminnassa. Sisäisessä toiminnassa ohjelmointirajapinnan avulla voidaan yhdistää haluttua analytiikkaa, jolloin jokaista analytiikkalukua ei tarvitse hakea Analyticsin verkkosovelluksen kautta. Mahdollisesti myös asiakkaille tai verkkosivun vierailijoille on mahdollista näyttää Analyticsiin perustuvaa dataa. Esimerkiksi sivuston katselukerrat voidaan näyttää sivustolla vierailevalle henkilölle. Sivustolta voidaan näyttää myös tarkempaa dataa, kuten lähteet tai vierailijoiden maantieteellinen sijainti. Mahdollisuuksia on hyvin paljon, eikä niitä kaikkia ole lueteltu tässä insinööriyössä.

Itselleni työ opetti verkkoanalytiikan käytöstä ja merkityksestä. Ohjelmointirajapintoihin tutustuminen ja sen testaaminen vaativat uuden opettelua ja kehittivät jo aikaisemmin opittua tietoa ohjelmointirajapintojen käytöstä. Projekti käsitti myös paljon HTML- ja CSS-tekniikoiden käyttöä, jotka olivat tuttuja jo ennestään. Kuitenkin niiden käyttö opettaa aina lisää ja käytössä tulee varmemmaksi. Käytössä oli myös uusia kirjastoja, kuten Data Tables -kirjasto.

Kokonaisuudessa työ oli onnistunut ja saavutti sille asetetut tavoitteet. Sovellusta ei ole julkistettu vielä julkiseen käyttöön, joten tuloksia oikeilta käyttäjiltä ei ole vielä saatavilla. Sisäisessä testauksessa sovellus on toiminut hyvin ja suurilta ongelmilta on vältytty. Tulevaisuudessa julkistuksen jälkeen sovelluksen suorituskykyä ja ominaisuuksia tarkkailaan, kunnes kaikki näyttää toimivan halutulla tavalla. Tämän jälkeen tarkkailua myös jatketaan ja mahdollisia parannuksia tehdään käyttäjäkokemuksia silmällä pitäen. Uusia ominaisuuksia mahdollisesti suunnitellaan.

Lähteet

- 1 Clifton, Brian. 2012. Advanced Web Metrics with Google Analytics. Third Edition. Canada: Sybex.
- 2 Khosrow-Pour, Mehdi. 2015. Encyclopedia of Information Science and Technology. Third Edition. Kappale: 756. IGI Global.
- 3 Vieraslista ja kävijälaskuri. Verkkodokumentti. Herdifier Oy. <<https://irc-galleria.net/info>>. Luettu 17.8.2016.
- 4 Bruce, Peter C. 2014. Introductory Statistics And Analytics. Wiley.
- 5 Saxena, Rahul; Srinivasan, Anand. 2013. Business Analytics. Dordrecht Springer.
- 6 Kaushik, Avinash. 2010. Web Analytics 2.0. Wiley.
- 7 Big data -määritelmiä. Verkkodokumentti. <<http://www.bigdata.fi/big-data-maari-telma>>. Luettu 13.9.2016.
- 8 Sathi, Arvind. 2014. Engaging customers using big data: how marketing analytics are transforming business. Palgrave Macmillan.
- 9 Schmarzo, Bill. 2013. Big Data: Understanding How Data Powers Big Business. Wiley.
- 10 Salo, Immo. 2013. Big data tiedon vallankumous. Docendo.
- 11 As Google Analytics Turns 10, We Ask: How Many Websites Use It? Verkkodokumentti. Matt McGee. <<http://marketingland.com/as-google-analytics-turns-10-we-ask-how-many-websites-use-it-151892>>. Luettu 17.8.2016.
- 12 Google Analytics Usage Statistics. Verkkodokumentti. BuiltWith® Pty Ltd. <<http://trends.builtwith.com/analytics/Google-Analytics>>. Luettu 17.8.2016.
- 13 Top Analytics Technologies. Verkkodokumentti. SimilarTech Ltd. <<https://www.similartech.com/categories/analytics>>. Luettu 17.8.2016.
- 14 Tracking Code Overview. Verkkodokumentti. Google Inc. <<https://developers.google.com/analytics/resources/concepts/gaConceptsTrackingOverview>>. Luettu 18.8.2016.
- 15 JPEG, GIF, or PNG? Image Filetypes Explained and Tested. Verkkodokumentti. Dann Albright. <<http://www.makeuseof.com/tag/jpeg-gif-png-image-filetypes-explained-tested/>>. Luettu 10.9.2016.

- 16 Overview. Verkkodokumentti. Google Inc. <<https://developers.google.com/analytics/devguides/reporting/core/v4/>>. Luettu 21.8.2016.
- 17 Dimensions & Metrics Explorer. Verkkodokumentti. Google Inc. <<https://developers.google.com/analytics/devguides/reporting/core/dimsmets>>. Luettu 21.8.2016.
- 18 Query Explorer. Verkkodokumentti. Google Inc. <<https://ga-dev-tools.appspot.com/query-explorer/>>. Luettu 21.8.2016.
- 19 Overview. Verkkodokumentti. Google Inc. <<https://developers.google.com/analytics/devguides/reporting/embed/v1/>>. Luettu 21.8.2016.
- 20 Embed API Server-side Authorization. Verkkodokumentti. Google Inc. <<https://ga-dev-tools.appspot.com/embed-api/server-side-authorization/>>. Luettu 21.8.2016.
- 21 What Is The Multi-Channel Funnels Reporting API – Overview. Verkkodokumentti. Google Inc. <<https://developers.google.com/analytics/devguides/reporting/mcf/v3/>>. Luettu 26.8.2016.
- 22 API Limits and Quotas. Verkkodokumentti. Google Inc. <<https://developers.google.com/analytics/devguides/reporting/core/v4/limits-quotas>>. Luettu 27.8.2016.
- 23 About sampled data. Verkkodokumentti. Google Inc. <<https://support.google.com/analytics/answer/1042498?hl=en>>. Luettu 27.8.2016.
- 24 Query Explorer. Verkkodokumentti. Google Inc. <<https://ga-dev-tools.appspot.com/query-explorer/>>. Luettu 21.8.2016.
- 25 Google Releases Analytics 360 Suite – First Impressions. Verkkodokumentti. Rich McPharlin. <<http://blog.performics.com/google-releases-analytics-360-suite-first-impressions-pn/>>. Luettu 28.8.2016.
- 26 Google APIs Terms of Service. Verkkodokumentti. Google Inc. <<https://developers.google.com/terms/>>. Luettu 19.9.2016.
- 27 Safeguarding your data. Verkkodokumentti. Google Inc. <<https://support.google.com/analytics/answer/6004245?hl=en>>. Luettu 19.9.2016.
- 28 Job Posts. Verkkodokumentti. LinkedIn. <<https://business.linkedin.com/talent-solutions/post-jobs#!>>. Luettu 21.9.2016.
- 29 Consumer Insights Manager – Healthcare. Verkkodokumentti. LinkedIn. <<https://www.linkedin.com/jobs2/view/206634652>>. Luettu 21.9.2016.

- 30 Google Analytics - All API Changelogs. Verkkodokumentti. Google Inc. <<https://developers.google.com/analytics/devguides/changelog>>. Luettu 19.9.2016.
- 31 Data Deprecation Policy for Google Analytics APIs. Verkkodokumentti. Google Inc. <<https://developers.google.com/analytics/devguides/reporting/core/v3/data-deprecation>>. Luettu 19.9.2016.
- 32 DataTables Table plug-in for jQuery. Verkkodokumentti. SpryMedia Ltd. <<https://datatables.net/>>. Luettu 28.8.2016.
- 33 Visualization: Area Chart. Verkkodokumentti. Google Inc. <<https://developers.google.com/chart/interactive/docs/gallery/areachart>>. Luettu 9.9.2016.
- 34 Error Responses. Verkkodokumentti. Google Inc. <<https://developers.google.com/analytics/devguides/reporting/core/v3/coreErrors>>. Luettu 11.9.2016.