



TAMPEREEN TEKNILLINEN YLIOPISTO

**TEEMO TEBEST
VERKKOPALVELUN KÄYTÖN SEURANTA JA
SEURANTATIEDON VISUALISOINTI**

Diplomityö

Tarkastajat: Ossi Nykänen &
Jukka Huhtamäki

Tarkastaja ja aihe hyväksytty tieto-
ja sähkötekniikan tiedekuntaneuvoston
kokouksessa 3. maaliskuuta 2010

Tiivistelmä

TAMPEREEN TEKNILLINEN YLIOPISTO

Tietotekniikan koulutusohjelma

TEBEST, TEEMO ANTON : Verkkopalvelun käytön seuranta ja seurantatiedon visualisointi

Diplomityö, 100 sivua

Kesäkuu 2010

Pääaine: Hypermedia

Tarkastajat: erikoistutkija Ossi Nykänen & tutkija Jukka Huhtamäki

Avainsanat: Erimenu.fi, verkkopalvelut, käytön seuranta, tiedon visualisointi

Verkkopalvelujen käytön seuraaminen on ajankohtainen ja kiinnostava tutkimuksen kohde, jota käsitellään kirjallisuudessa. Tutkimustyön tavoitteena oli toteuttaa verkkopalveluun käyttöä seuraava järjestelmä. Järjestelmän tehtävä oli tuottaa kerätyn tiedon perusteella verkkopalvelun ylläpitäjille, kehittäjille ja mahdollisesti muille sidosryhmille ajantasaista tietoa verkkopalvelun käytöstä. Työn teoriaosuudessa käytettiin pohjana verkkopalvelujen käytön seuraamiseen liittyvää kirjallisuutta. Työn empiirinen osuus toteutettiin Erimenu.fi-palveluun, joka on erityisruokavalion omaavan henkilön ruokavaliota tukeva verkkopalvelu. Erimenu.fi-verkkopalvelulla on päivittäin noin 500 käyttäjää, joka tekee tutkimuksessa käsiteltävästä tiedosta todellista.

Tutkimustyössä esiteltiin kolme keskeisintä teknologista käytön seurantakeinoa, jotka olivat käytön seuranta 1) verkkopalvelun HTTP-palvelimella, 2) käyttäjän asiakasohjelmassa ja 3) sovellustasolla integroituna verkkopalveluun. Nämä kolme teknologista tapaa pystyttiin erittelemään verkkopalvelujen käytön seurantaan liittyvästä kirjallisuudesta. Seurannan mahdollistavien teknologisten ratkaisujen jälkeen selvitettiin millä tavoin valitulla seurantateknologialla kerätty tieto on mahdollista visualisoida. Työssä toteutettiin kerätyn teorian pohjalta verkkopalvelun käyttöä sovellustasolla seuraava käytön seurantajärjestelmä ja kerättyä tietoa visualisoivia näkymiä. Tutkimuksessa käytettiin kvantitatiivista tutkimusotetta, koska työssä toteutetun järjestelmän oli tarkoitus kerätä seurantatietoa automatisoidusti, joka rajoitti mahdollisuuksia kvalitatiivisen tiedon keräämiseen. Kvalitatiivinen tutkimusote otettiin huomioon työn teoriaosassa. Toteutustyö tehtiin Erimenu.fi-verkkopalveluun hyödyntämällä Python-ohjelmointikieleen perustuvaa Django-kehitysympäristöä. Toteutetun järjestelmän avulla mahdollistettiin etenkin verkkopalvelun tietosisältöihin liittyvän tiedon seuranta.

Tehdyn käytön seurantajärjestelmän perusteella todettiin, että verkkopalveluun integroituvalla seurantajärjestelmällä on mahdollista seurata verkkopalvelun käyttöä tietyiltä osin huomattavasti laajemmin kuin muilla ratkaisumalleilla. Tutkimuksessa havaittiin, että verkkopalveluun liittyvää sisäistä tietoa ei ole mahdollista seurata ellei seurantajärjestelmää sisällytetä osaksi verkkopalvelua. Työn päätteeksi ratkaisumalleja arvioitiin monikriteerisen päätöksentekomallin avulla, jonka tueksi valittiin kirjallisuuden perusteella verkkopalvelun käytön seurantajärjestelmiä kuvaavia ominaisuuksia.

Abstract

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Information Technology

TEBEST, TEEMO ANTON : Web service monitoring and
visualization of surveillance data

Master of Science Thesis, 100 pages

June 2010

Major: Hypermedia

Examiners: senior researcher Ossi Nykänen & researcher Jukka Huhtamäki

Keywords: Erimenu.fi, web services web monitoring, data visualization

The monitoring of the use of web services is a current and interesting subject of a study which is dealt with in literature. The objective of the research work was to implement a monitoring system to a web service. System's task was to produce real-time information about the use of the web service on the basis of the collected information to the administrators, system developers and other reference groups. Literature which is related to the monitoring of the use of web services was used in the theory share of the work. The empiric share of the work was carried out in Erimenu.fi which is a web service that supports users who has a special diet. Erimenu.fi service has about 500 daily users which makes the information dealt with in the study qualitative.

The research work presented three most central technological web service monitoring methods that were monitoring the usage 1) on the HTTP server of the web service, 2) in the user's client program and 3) at the application level integrated into the web service. These three technological ways were possible to identify from web service related literature. After clarifying the technological solutions which make the monitoring possible it was defined how it is possible to visualize the information that was collected with the chosen monitoring technology. The system which monitors the use of the web service at the application level and the views which visualize the collected information were implemented in the work based on the collected theory. In the study a quantitative study extract was used because it was intended for the system that has been carried out in the work to collect the surveillance data automatically which restricted opportunities to collect qualitative information. The qualitative study extract was taken into consideration in the theory part of the work. The implementation to Erimenu.fi was done by utilising the Django software framework which is based on Python scripting language. With the help of the executed system the monitoring of especially the information which is related to the data contents of the web service was made possible.

On the basis of the created monitoring system it was stated that on some parts it is possible to monitor web services more widely with a system that is integrated to the service than with other available monitoring solutions. In the study it was noticed that it is not possible to monitor the internal information which is related to the web service if the monitoring system is not included as a part of the web service. The solutions were analysed with multi-criteria decision analysis method and attributes for the analysis were chosen from the web service related literature.

Alkusanat

Diplomityön taustana on vuonna 2008 aloitettu Aatu-projekti, jossa Tampereen teknillisen yliopiston hypermedialaboratoriossa toteutettiin muun muassa Raha-automaattiyhdistyksen varoilla Erimenu.fi-niminen käyttäjän erityisruokavaliota tukeva verkkopalvelu. Henkilökohtainen taustani on, että tulin töihin hypermedialaboratorioon matematiikan laitokselle vuoden 2009 alussa. Kevään 2009 aikana toimin verkkoyhteisöjen kehitysryhmässä, jossa olin mukana useiden verkkopalvelujen toteuttamisessa. Näiden projektien ja aikaisemman kokemukseni kautta diplomityöni pääasiallinen ohjaaja Jukka Huhtamäki havaitsi kiinnostukseni verkkopalvelujen käytön näkyväksi tekemiseen ja Erimenu.fi-verkkopalvelussa oli tarve järjestelmälle, jonka avulla voitaisiin seurata miten palvelua käytetään.

Diplomityö aloitettiin kehitystyön osalta kesällä 2009. Erimenu.fi-verkkopalvelussa käytetty kehitysympäristö Django ja ohjelmointikieli Python olivat minulle entuudestaan tuntemattomia, mutta aikaisemmassa toteutustöissä mukana olleet Pasi Häkkinen sekä Antti Mäki ohjeistivat ja auttoivat minua loistavasti aina tarvitessani apua. Diplomityön kirjoitusprosessin aloitin aktiivisesti syksyn 2009 aikana. Toteutuksen osalta työssä esitellyt tulokset saatiin valmiiksi vuoden 2010 alussa. Kirjoitustyö jatkui aivan diplomityön valmistumiseen asti kesäkuuhun 2010.

Kiitokset tämän työn valmistumisesta haluan esittää ensinnäkin tyttöystävälleni Anulle, joka on tukenut minua koko työn ajan. Toiseksi haluan kiittää koko hypermedialaboratorion ja matematiikan laitoksen väkeä, joista erityiskiitos kuuluu diplomityöni ohjaajille Ossi Nykänen ja Jukka Huhtamäki. Kiitos kuuluu myös projektin rahoittaneelle Raha-automaattiyhdistykselle sekä Pirkanmaan Allergia- ja Astmayhdistys ry:lle, joita ilman tämän diplomityön taustalla olevaa Erimenu.fi-verkkopalvelua ei olisi olemassa. Kiitos Tatu Anttoselle diplomityön läpi lukemisesta ja kommentoimisesta. Viimeisimpänä mutta ei todellakaan vähäisimpänä haluan osoittaa kiitokseni rakkaalle äidilleni.

26.5.2010

Opiskelijankatu 4 A 16

33720 TAMPERE, SUOMI

TEEMO TEBEST

Sisällys

1	Johdanto	1
1.1	Tutkimusongelma	1
1.2	Työn tavoitteet ja rajaukset	2
1.3	Työn taustat	3
1.4	Työn rakenne	3
2	Verkkopalvelun käytön seuranta	4
2.1	Käytön seurannan historia	5
2.2	Käytön seurantatieto	8
2.2.1	Kvantitatiivinen tieto verkkopalveluissa	9
2.2.2	Seurantatiedon merkitys	10
2.3	Asiakas-palvelin -malli	10
2.3.1	Asiakasohjelmat	11
2.3.2	Palvelinohjelmat	12
2.3.3	Asiakas-palvelin -malli verkkopalveluissa	13
2.4	Palvelintason käytön seuranta	14
2.4.1	HTTP-metodit	15
2.4.2	HTTP-vastaukset	16
2.4.3	Apache	17
2.5	Asiakasohjelman käytön seuranta	19
2.6	Käytön seurantajärjestelmä	20
2.6.1	AWStats	22
2.6.2	Google Analytics	23
2.7	Sovellustason käytön seuranta	24
2.7.1	Tietokannan rooli sovellustason käytön seurannassa	24
2.7.2	Seurattavien toimintojen tunnistaminen sovellustasolla	25
2.7.3	Sovelluksen käyttäjätunnuksen liittäminen asiakasohjelmaan	25
2.7.4	Käytön seurannan tehokkuuskysymykset	26
2.8	Käytön seurannan vaikutukset verkkopalvelun käyttäjään	26
3	Seurantajärjestelmän toteutus verkkopalvelussa	29
3.1	Verkkopalveluiden arkkitehtuuri	30
3.2	Sivulatausten seuranta	32
3.2.1	Sivulatausten suodattaminen	33
3.2.2	Verkon ulkopuolella tapahtuvan käytön seuraaminen	34
3.2.3	Vaihtoehtoiset sisällön lukutavat	35
3.3	Seurattavien tietojen määrittely	35
3.3.1	Neljä suurta kysymystä	36
3.3.2	Kaistan käytön seuranta	38
3.3.3	Palvelinkuorman seuranta	39
3.3.4	Asiakaslaitteiden monikäyttö	39
3.3.5	Mobiililaitteiden seuranta	40
3.3.6	Paikkatiedon seuranta	41
3.3.7	Käyttäjän ympäristön seuranta	41
3.4	Tapahtumien seuranta	42

3.4.1	Tapahtumasarjojen seuranta	42
3.5	Yhteisöllisen käytön seuranta	43
3.5.1	Yhteisöllisyys verkkopalvelun käytön motivaattorina	45
3.6	Seurantatiedon tallennus	47
3.6.1	Tietokantatallennus	47
3.6.2	Tiedostotallennus	48
3.7	Käytön seurantajärjestelmän käyttöönotto	48
3.7.1	Seurantajärjestelmän toteuttaminen uuteen verkkopalveluun	48
3.7.2	Seurantajärjestelmän sisällyttäminen verkkopalveluun	49
3.8	Käytön seurantatiedon vienti ja tuonti	49
3.8.1	Seurantatiedon vienti järjestelmästä	50
3.8.2	Seurantatiedon tuonti järjestelmään	51
4	Verkkopalvelun käytön visualisointi	53
4.1	Käytön seurantatiedon keräysprosessi	53
4.2	Visualisointi	54
4.2.1	Tiedon visualisointi ja tiedon analysointi	55
4.2.2	Visualisointitekniikat Internetissä	55
4.2.3	Kuvaajat verkkopalvelujen visualisoinnissa	56
4.2.4	Tunnusluvut verkkopalvelujen visualisoinnissa	57
4.2.5	Käyttäjien segmentointi	57
4.2.6	Geologisen paikan visualisointi	58
4.2.7	Haasteet visualisointien toteutuksessa verkkopalveluun	59
4.3	Visualisoinnin periaatteet	59
4.3.1	Luettavuus	59
4.3.2	Asettelu	60
4.3.3	Asteikko	61
4.3.4	Luotettavuus	62
4.3.5	Määrä vs. kertymä	62
4.4	Visualisointityökalut	63
4.4.1	SIMILE Widgets	63
4.4.2	Ext JS	64
4.4.3	Taulukkolaskentaohjelmistot	65
4.4.4	Google Maps	66
4.4.5	Visualisointiympäristön valinta	67
4.5	Visualisoinnin jakelu	68
4.5.1	Visualisoinnin toteutus verkkopalvelussa	68
4.5.2	Visualisointi verkkopalvelun ulkopuolella	68
5	Käytön havainnointi: Erimenu.fi	69
5.1	Erimenu.fi	69
5.2	Sivulatausten seuranta Erimenu.fi:ssä	71
5.3	Erimenu.fi-verkkopalvelusta tuotettu lokitieto	76
5.4	Muutoksen havainnointi	78
5.4.1	Havainnointi kaavioiden ja tunnuslukujen avulla	78
5.4.2	Havainnointi karttavisualisointien avulla	82
5.5	Muutoksen selittäminen	83

6 Tulosten arviointi	85
6.1 Verrattuna palvelinpään käytön seuraimeen	88
6.2 Verrattuna asiakasohjelmassa toimivaan käytön seuraimeen	89
7 Yhteenveto	91
Lähteet	94

Kuvat

2.1	Internetissä asiakkaita (client system) edustavat selainohjelmistot. Tietovarastona (data store) toimii HTTP-palvelin. Yhteydet luodaan asiakkaiden ja tietovaraston välille. (<i>Distributed Application Architecture</i> , 2007, s. 129)	11
2.2	Palvelimen palauttamien HTTP-vastauksien määriä on mahdollista seurata verkkopalveluissa.	17
2.3	AWStatsin tuottama visualisointi verkkopalvelun käytöstä kuukausittain.	22
2.4	Esimerkki sivulatausten määrää kuukauden ajalta kuvaavasta viivadiagrammista Google Analytic -palvelussa.	23
3.1	MVC-arkkitehtuurin rakenne ja osien väliset yhteydet.	31
3.2	Käytön seurantatietoa voidaan käyttää rikastamaan verkkopalvelun käyttäjän käyttökokemusta, kuten nähdään tässä keskustelualan säikeen kohdalla.	32
3.3	Sosiaalisen vuorovaikutuksen seuranta verkkopalveluissa. (Silius et al., 2009, s. 5)	45
3.4	Erimenu.fi:ssä sosiaalista käyttöä tuodaan käyttäjän näkyville avain-sana ja oma lista -merkintöjen osalta. Käyttäjien yksityisyyteen liittyvät tiedot on sumennettu.	46
4.1	Visualisointilähtöinen malli tietojoukon esittämiseen. (Ware, 2004, s. 12)	53
4.2	Esimerkki pystyyn asetellusta pylväsdiagrammista. Käyttäjien yksityisyyteen liittyvät tiedot on sumennettu.	56
4.3	Esimerkki vaakatasossa olevasta pylväsdiagrammista. Visualisoinnissa pienimmät arvot ovat kuvaajateknisistä syistä skaalattu samaan leveyteen. Käyttäjien yksityisyyteen liittyvät tiedot on sumennettu.	56
4.4	Kuvaajia esitettäessä on hyvä tarjota käyttäjälle mahdollisimman paljon tietoa tiedosta.	60
4.5	Kun visualisoinnit esitetään allekkain samalla aikajanalla on kuvaajien arvojen vertailu helppoa samoilta x-akselin arvoilla.	61
4.6	Timeplot on SIMILE Projectin luoma avoimeen lähdekoodiin perustuva visualisointityökalu. (<i>Timeplot; Web Widget for Plotting Time Series</i> , s.a.)	64
4.7	Ext JS on JavaScript kirjasto, jolla voidaan tehdä hyvin monipuolisia ja rikkaita verkkopalveluja (<i>Ext JS; Charts</i> , s.a.)	65
4.8	Google Mapsilla voidaan luoda sijaintitietojen perusteella esimerkiksi visualisointi halutun artistin tulevista keikoista.	67
5.1	Lähipiiriin on mahdollista lisätä muita Erimenu.fi-verkkopalvelun käyttäjiä. Käyttäjien yksityisyyteen liittyvät tiedot on sumennettu.	71
5.2	Sivulataustiedon tallentavan tietokantataulun rakenne.	75
5.3	Ainesosa-tietosisällön tallentavan tietokantataulun rakenne.	75
5.4	Lokitiedon viennin toteuttava lomake Erimenu.fi:ssä.	77
5.5	Kuukausittainen uusien yksityiskäyttäjien määrä Erimenu.fi-verkkopalvelussa.	79

5.6	Uusien Erimenu.fi-yksityiskäyttäjien määrä päivittäin joulukuussa 2009.	79
5.7	Uusien Erimenu.fi-yksityiskäyttäjien määrä päivittäin joulukuussa 2008.	79
5.8	Yksityiskäyttäjätilien kertymä kuukausittain Erimenu.fi-verkkopalvelussa.	80
5.9	Erimenu.fi-verkkopalvelussa luettavissa oleva tiedote.	81
5.10	Erimenu.fi:ssä päivittäin kirjautuneiden käyttäjien määrä.	81
5.11	Erimenu.fi:ssä päivittäin vierailevien IP-osoitteiden määrä.	82
5.12	Erimenu.fi-verkkopalvelussa tehtyjen Lähipiiri-yhteyksien kertymä. . .	82
5.13	Vasemmalla kuvataan Erimenu.fi-palveluun rekisteröityneet käyttäjät vuoteen 2009 mennessä ja oikealla vuoteen 2010 mennessä.	83
6.1	Erimenu.fi-palvelussa tehty amarantti-päivitys näkyy selkeästi HTTP-palvelimen lokitietoa jäsentävässä AWStats-ohjelmassa.	89
6.2	Päivittäisten käyttäjien määrä Erimenu.fi-verkkopalvelussa aikajak-solla tammi-helmikuu 2010 Google Analytics -palvelussa nähtynä. . .	90

Taulukot

2.1	HTTP-metodit ja niiden käyttötarkoitukset. (<i>Hypertext Transfer Protocol</i> , 1999, ss. 51-56; Jacobs, 2004)	15
2.2	Esimerkki GET-metodilla välitetyistä HTTP-otsikkotiedoista.	16
2.3	Jäsennetty Apachen keräämä lokitieto.	19
5.1	Sivulatauksen yhteydessä tallennettavat reseptiin liittyvät tiedot.	72
5.2	Sivulatauksen yhteydessä tallennettavat tuottajaan liittyvät tiedot.	72
5.3	Sivulatauksen yhteydessä tallennettavat tuotteeseen liittyvät tiedot.	73
5.4	Sivulatauksen yhteydessä tallennettavat avainsanaan liittyvät tiedot.	73
5.5	Sivulatauksen yhteydessä tallennettavat lähipiiriin liittyvät tiedot.	73
5.6	Sivulatauksen yhteydessä tallennettavat ruokavalioon liittyvät tiedot.	73
5.7	Sivulatauksen yhteydessä tallennettavat käyttäjäprofiliin liittyvät tiedot.	74
5.8	Sivulatauksesta tallennettavat yleiset ominaisuudet.	74
6.1	Käytön seurantateknologioiden vertailutaulukko.	87
6.2	Käytön seurantajärjestelmien vertailutaulukko.	88

Lyhenteet ja merkinnät

AJAX Asynchronous JavaScript And XML

API Application Programming Interface

CLF Common Log Format

CSS Cascading Style Sheets

CSV Comma-Separated Values

DHTML Dynamic HTML

DNS Domain Name Service

ELF Extended Log Format

FTP File Transfer Protocol

GPS Global Positioning System

GSM Global System for Mobile Communications

HTML Hypertext Markup Language

HTTP Hypertext Transfer Protocol

IP Internet Protocol

JS JavaScript

LAN Local Area Network

PDF Portable Document Format

PERL Practical Extraction and Report Language

PHP Hypertext Preprocessor

RSS Really Simple Syndication

SNA Social Network Analysis

SQL Structured Query Language

TCP Transmission Control Protocol

URI Uniform Resource Identifier

URL Uniform Resource Locator

WLAN Wireless Local Area Network

WWW World Wide Web

XML Extensible Markup Language

1 Johdanto

Tässä diplomityössä luodaan katsaus verkkopalvelujen (web services) käytön seurannan (web monitoring) teknologioihin ja kerätyn tiedon visualisointikeinoihin. Työstä lukija saa kuvan verkkopalvelujen käytön seuraamisen nykytilasta ja minkä tyyppisiä asioita käytön seurantajärjestelmän toteutuksessa ja käyttöönotossa tulisi ottaa huomioon. Edellä mainittuja asioita käsitellään yleisluontoisesti, mutta teorian tukena esitetään esimerkkejä käytännön järjestelmistä. Tutkimustuloksina esitellään Erimenu.fi-verkkopalveluun kehitetyn käytön seurantajärjestelmän toteutustyön tulokset sekä kerätystä tiedosta koostetut visualisoinnit. Toteutettavan järjestelmän nähdään ensisijaisesti tulevan käyttöön järjestelmän ylläpitäjille ja kehittäjille. Järjestelmän avulla on kuitenkin mahdollista palvella verkkopalvelun tavallisia käyttäjiä.

Tutkimuksen kannalta Erimenu.fi-verkkopalveluun keskittyminen johtaa siihen, että tutkimus ja sen aineisto ovat tapaustutkimusta. Tarkemmin tutkimus toteutetaan yhdessä tutkimuskohteessa, jossa aineistoa kerätään kahtena ajankohtana tietyn kriittisen tapahtuman molemmin puolin (“pre-post case studies”). Kriittisenä tapahtumana toimii Erimenu.fi-palveluun toteuttava Lähipiiri-toiminnallisuus. Tapaustutkimus on luonteeltaan kertaluonteista ja vaikeasti yleistettävää, koska se keskittyy yhden tietyn ilmiön selittämiseen. (Garson, 2008; Järvinen & Järvinen, 2004, ss. 75-82)

1.1 Tutkimusongelma

Työssä on tehtävänä toteuttaa olemassa olevaan verkkopalveluun kattava käytön seurantajärjestelmä ja luoda kerätystä tiedosta havainnollisia visualisointeja. Tutkimuskysymys on: “miten pystytään havaitsemaan muutos verkkopalvelun käytössä?”. Kysymyksen tukena työssä vastataan kysymyksiin “millä teknologiaratkaisuilla verkkopalveluiden seuraaminen on mahdollista?”, “miten kerätty seurantatieto esitetään, että sen havaitseminen on mahdollisimman havainnollista?” ja “millä tavoin toteutettavan seurantajärjestelmä keräämä tieto on kattavampaa kuin vaihtoehtoisilla teknologioilla ja järjestelmillä kerättävä tieto?”. Työssä vastataan “onko mahdollista toteuttaa nykyisiä yleisiä käytön seurantajärjestelmiä kattavuudeltaan vastaava järjestelmä?”. Verkkopalvelun käytössä havaittavan muutoksen osalta pyritään vastaamaan kysymyksiin: “pystytäänkö verkkopalvelun käytössä havaitsemaan tapahtuva muutos kun järjestelmään toteutetaan uusi ominaisuus tai päivitys?” ja “miten tämä muutos voidaan yksilöidä verkkopalvelun yksittäisiin ominaisuuksiin?”. Toisaalta tutkitaan “voidaanko seurata uusien ominaisuuksien käyttöönoton suosiota?”. Tutkimuksessa pyritään vastaamaan myös siihen “mitä tietoa verkkopalvelun käyttäjistä pystytään automaattisesti ilman käyttäjän interaktiota keräämään?”. Työssä

sivutaan kysymystä: “mikä käytön seurantateknologia on kustannuksiltaan paras ratkaisu?”.

1.2 Työn tavoitteet ja rajaukset

Työn tavoitteena on kuvata sellaisen käytön seurantajärjestelmän ja seurantatiedon visualisointien kokonaisuutta, jossa verkkopalvelun käyttöä pystytään kuvaamaan paremmin kuin pelkästään sarjana peräkkäisiä sivulatauksia (clickstream). Tavoite on merkittävä, koska verkkopalvelujen käytöstä ei saatavilla olevilla teknologioilla saada sellaista käyttötietoa, jonka avulla pystyttäisiin kuvaamaan käyttäjiä verkkopalveluun sitoen. (Vrt. Kaushik, 2010, ss. 5-8)

Työn tuloksina toteutettavien visualisointien on tarkoitus mahdollistaa verkkopalvelussa tapahtuva käytön muutoksen havainnointi. Muutosta verkkopalvelun käyttöön pyritään tuottamaan verkkopalveluun toteutettavilla uusilla ominaisuuksilla. Visualisointien taustalle verkkopalveluun on tarkoitus toteuttaa erilaisia käytön seurantaa toteuttavia komponentteja. Visualisointien lukijalle on tarkoituksena pyrkiä esittämään erilaisia aikaikkunoita, joissa tietoa voidaan tarkastella. Tarkastelijalle on tarkoituksena antaa vapaus mitä tietoa hän tutkii ja minkälaisella aikajanalla hän haluaa tehdä havaintonsa.

Muutoksen syiden selittäminen rajataan tämän työn ulkopuolelle. Työn tarkoituksena ei ole analysoida kerättyä käytön seurantatietoa. Analysoinnilla tarkoitetaan, että ei yritetä vastata kysymykseen “miksi”. Työn tavoitteena ei ole visualisointien avulla pyrkiä selittämään verkkopalvelun käytössä tapahtuvaa muutosta, vaan kehittää järjestelmä, jonka avulla ilmiöiden ja niissä tapahtuvien muutosten havaitseminen on mahdollista. Huomioitavaa kuitenkin on, että tietynlaista analyysiä tiedolle joudutaan tekemään, kun pohditaan minkä tyyppisiä visualisointeja sen perusteella on järkevää tuottaa. Työssä pyritään toteuttamaan työkaluista sellaisia, että tiedon havaitsija pystyy tulkitsemaan koostettua tietoa mahdollisimman autenttisesti.

Seurantajärjestelmän toteutuksessa vältetään toistamasta jo Erimenu.fi-verkkopalvelussa käytössä olevissa visualisointipalveluissa olevia ominaisuuksia. Työn on tarkoituksena luoda verkkopalvelun käytöstä etenkin sellaista tietoa, jota ei ole vielä saatavilla.

Diplomityössä ei tutkita tarjolla olevia kaupallisia käytön seurantajärjestelmiä. Rajaus tehdään, koska työssä pyritään kuvaamaan enemmän käytön seurannan taustalla olevaa teknologiaa. Toteutusteknisesti projektissa valittiin suunnaksi oman käytön seurantatoteutuksen tekeminen, joka osaltaan vaikuttaa siihen miksi tarjolla olevia kaupallisia järjestelmiä ei tässä työssä tuoda esiin.

1.3 Työn taustat

Työssä käsiteltävää käytön seurannan aihealuetta sivutaan aikaisemmissa julkaisuissa, joissa verkkopalvelujen käyttöä on tutkittu etenkin SNA-malleihin (Social Network Analytisis) perustuen. Aikaisempia julkaisuja, joissa olen ollut mukana ovat (Silius et al., 2009) ja (Silius et al., 2010). Tutkimustyö on tehty Tampereen teknillisen yliopiston hypermedialaboratoriossa, jossa olen toiminut 2009 vuoden alusta tutkimusapulaisena. Näissä tutkimuksissa on keskitytty verkkopalvelujen käyttäjien välisiin suhteisiin ja analysoitu niitä etenkin matemaattisten mallien avulla, kun tässä diplomityössä tarkoituksena on tutkia verkkopalvelun käytön seurantajärjestelmän kokonaisvaltaista toteutusta. Oma roolini tutkimuksissa ollut toimia tutkimusapulaisen roolissa ja liittynyt tutkimusympäristöjen tekniseen kehitystyöhön, joten diplomityön aihealue on looginen jatke aikaisemmin tekemälleni tutkimustyölle. Aliluvussa 3.5 käydään läpi aikaisemman tutkimuksen tämän työn kannalta keskeisimmät tulokset. (Vrt. Krebs, 2008)

1.4 Työn rakenne

Diplomityön alussa käsitellään verkkopalvelujen käytön seuraamiseen liittyvää teoriaa. Työssä aikaan saadut tulokset esitellään työn lopussa. Luvussa 2 annetaan lukijalle kokonaiskuva verkkopalvelujen käytön seuraamisesta. Luvussa 3 kuvataan millä tavoin käytön seurantajärjestelmän toteuttaminen on mahdollista suoraan verkkopalvelun yhteyteen. Luvussa 4 esitellään verkkopalvelujen käytön seurantatiedon visualisointiin liittyvää teoriaa.

Työssä toteutetut käytön seurantakomponentit ja visualisoinnit käydään läpi luvussa 5. Luvussa 6 arvioidaan luvussa 5 toteutettua käytön seurantajärjestelmää muihin verkkopalvelun käytön seuranta toteuttaviin järjestelmiin. Arviointimallina käytetään monikriteeristä päätöksentekoa. Luvussa 7 kootaan työn tulokset yhteen ja pohditaan työhön liittyvää mahdollista jatkotutkimuksen tarvetta.

2 Verkkopalvelun käytön seuranta

Verkkopalvelu (web service) on Internetissä toimiva järjestelmä, joka tarjoaa käyttäjille erilaisia toimintoja ja sisältöä. Käyttäjällä tarkoitetaan verkkopalvelua käyttävää henkilöä. Käyttäjä voi olla verkkopalvelussa satunnaisesti vieraileva henkilö tai toisaalta verkkopalvelua päivittäin hyödyntävä rekisteröitynyt käyttäjä. Verkkopalvelujen käytön seurannalla (web service monitoring) tarkoitetaan verkkopalvelun käyttäjien (user) toimintojen (action) seuraamista verkkopalvelussa.

Verkkopalvelujen käytön seurantatiedon keräämisen motivaationa voi olla halu selvittää verkkopalvelun käyttöön liittyviä tunnuslukuja (statistic). Verkkopalvelun käytön seurannasta kerättäviä tunnuslukuja ovat muun muassa verkkopalvelun kuukausittainen kävijämäärä tai verkkopalveluun rekisteröityneiden käyttäjien lukumäärä. On mahdollista, että käytön seurantatietoon on suoranainen tarve. Käytön seurannan avulla saatavalla tiedolla on mahdollista osoittaa organisaatioiden verkkopalveluissa olevia ongelmakohtia. Ongelmakohtat voivat vaikuttaa organisaation liiketoiminta edellytyksiin, joten käytön seurantatiedon kerääminen on tärkeää.

Käytön seurantatietoa voidaan hyödyntää verkkopalvelun kehitystyössä. Verkkopalveluun tuotavien uusien ominaisuuksien osalta käytön seurannasta saatava tieto on oleellinen osa sitä prosessia, jolla voidaan tutkia uusien ominaisuuksien toimivuutta. Käytön seurantatiedolla voidaan seurata miten nopeasti verkkopalvelun käyttäjät ottavat uudet ominaisuudet ja toiminnot käyttöönsä. Toisaalta käytön seurantatiedon tuella voidaan pyrkiä päättämään minkä tyyppisiä ominaisuuksia verkkopalveluun tulisi jatkossa kehittää. Verkkopalvelun käytön seurantatieto kasvattaa organisaatioissa olevan tiedon määrää ja lisääntynyttä tietoa voidaan hyödyntää organisaation päätöksenteossa.

Seuraamalla verkkopalvelun käyttöä pystytään paremmin käsittämään mitä verkkopalveluiden käyttäjät verkkopalveluissa tekevät. Internet-ympäristöissä kosketuspinta asiakkaan ja verkkopalvelua ylläpitävän organisaation välillä jää usein heikoksi ja “tuntuma asioiden tilaan” ei ole yhtä selkeä kuin reaali maailman prosesseissa. Ihmisten välisessä kommunikoinnissa tulkintaa voidaan tehdä aistien avulla, joka sähköisessä asiointissa ei ole mahdollista. Käytön seuraamisella tätä tunteen tunnetta ihmisten välisistä kontakteista pyritään tuomaan mukaan sähköiseen asiointiin. Verkkopalvelua ylläpitävän tahon on mahdollista oppia tuntemaan miten verkkopalvelua käytetään ja pystyä kehittämään palvelua suuntaan, joka tukee käyttöä paremmin.

Tarve verkkopalvelun käytön seuraamiselle on kasvanut myös Internetissä tapahtuvan mainonnan lisääntymisen vuoksi. Verkkopalvelua ylläpitävälle taholle on tarkoituksenmukaista pystyä todistamaan verkkopalvelussa mainostaville sidosryhmille ja muille toiminnan rahoittajille, että muun muassa heidän mainoksensa tule-

vat verkkopalvelussa nähdäiksi. Verkkopalvelun käytön seurantatietoa voidaan käyttää hyväksi pyrittäessä todistamaan oman organisaation suosiota ja tunnettavuutta. Jos esimerkiksi organisaatio on hakemassa toiminnalleen yhteistyökumppaneita, on organisaation verkkopalvelun päivittäinen kävijämäärä hyvä tunnusluku, jonka avulla voidaan osoittaa yhteistyökumppaneille, että mahdollinen tuleva yhteistyö on molemmin puolin kannattavaa. (Vrt. Kaushik, 2010, s. 1; vrt. Croll & Sean, 2009, s. 3)

Käytön seurantatiedon keräämiseen ja näkyväksi tekemiseen on olemassa tutkimuksellista kiinnostusta. Internet on ilmiönä vielä niin nuori ja alati kehittyvä, että sen määrittelyminen on erittäin vaikeaa. Alan ajankohtaisia ilmiöitä käsittelevä kirjallisuus sisältää käytännössä jo julkaisuhetkellään vanhentunutta tietoa. Tarve analyttiseen tutkimustietoon kehittyvässä Internet-ympäristössä on kuitenkin olemassa, koska ihmisten päivittäinen ajankäyttö Internetissä kasvaa jatkuvasti. (Vrt. *Internet Usage Statistics*, 2009; vrt. “Netissä roikutaan jo liki tunti päivässä”, 2010)

2.1 Käytön seurannan historia

Lokitiedon (log data) keräämisen ja sitä kautta käyttäjän toimien seurannan historia Internetissä ja verkkopalveluissa ulottuu niin kauas kuin HTTP-palvelimia on ollut olemassa. Kuten nykyään, 1990-luvun alussa Internetin alkaessa yleistyä, palvelimet käsitelivät niille tulleet pyynnöt ja kirjoittivat niistä merkinnän lokitiedostoon (log file). Sivulatauksista kerätyt ja ylös kirjatut tiedot eivät olleet kuitenkaan kattavia, eikä tiedolle oltu määritelty rakennetta, jonka avulla sen tehokas käsittely olisi ollut mahdollista. Sivulatauksiin liittyvät lokitiedot olivat kuitenkin luettavissa levyiltä ja niitä saatettiin satunnaisesti käyttää virheiden etsimiseen verkkopalvelusta. Kerättyyn lokitietoon oli usein pääsy vain organisaatioiden tietotekniikasta vastaavilla ihmisillä, eikä lokitiedon merkitystä organisaatiolle taloudellisista näkökulmista osattu siten nähdä.

Internetin yleistymisen myötä tarve lokitiedon käsittelemiselle kasvoi. Ensimmäisiä rakenteisia lokitietomalleja oli CLF (Common Log Format). CLF sisälsi tiedot yhteyden ottaneesta käyttäjästä ja kattavat pyyntöä koskeneet tiedot. Alla esimerkki CLF-muotoisesta lokitietomerkinästä, josta nähdään käyttäjän IP-osoite (Internet Protocol (IP) address), tapahtuman aikaleima, HTTP-metodi (HTTP-method), haettu resurssi ja HTTP-vastaus (HTTP-response).

```
127.0.0.1 - frank [10/Oct/2000:13:55:36 -0700]
"GET /apache_pb.gif HTTP/1.0" 200 2326
```

Internet-teknologioiden kehityksen myötä lokitietoon tuli mahdolliseksi kerätä tieto asiakasohjelmasta ja “referer”-tieto (viittaava sivu). Viittaavalla sivulla tarkoitetaan tietoa, joka kertoo miltä sivulta käyttäjä siirtyi kyseiselle sivulle. Vuonna 1996 julkistettiin CLF:stä laajennettu lokitietomalli ELF (Extended Log For-

mat). ELF-muotoon tallennettu lokitieto oli kuitenkin edelleen puutteellista, koska se ei sisältänyt juurikaan tietoa verkkopalvelussa vierailleesta käyttäjästä. ELF-muodossa kerätty tieto keskittyi HTTP-palvelimen tekniseen puoleen; mitä resursseja haettiin, mitkä verkko-osoitteet hakivat niitä, milloin niitä haettiin ja mikä vastaus näihin pyyntöihin palautettiin. ELF muotoon tallennetusta lokitiedosta oli mahdollista koostaa näkymiä, joiden avulla pystyttiin näyttämään mitä verkkopalvelussa tapahtui. Pääasiallisesti lokitietoa käytettiin kuitenkin edelleen vain verkkopalvelun ylläpidon toimesta virheiden etsimiseen. Alla esimerkki ELF muodossa olevasta lokitiedosta, josta voidaan lukea

- HTTP-palvelimen versio
- lokitietomuodon versio
- lokitiedoston luontiaika
- sivulataustapahtuman päivämäärä
- sivulataustapahtuman aika
- asiakkaan IP-osoite
- asiakkaan portti
- palvelimen IP-osoite
- palvelimen portti
- HTTP-metodi
- pyydetyn resurssin osoite
- HTTP-vastaus
- käyttäjän asiakasohjelman tiedot ja
- käyttäjän käyttöympäristön tiedot.

Huomioitavaa on, että ELF muotoon voidaan tallentaa edellä mainittujen lisäksi muita kenttiä. Suositut HTTP-palvelinjärjestelmät Apache¹ ja Microsoft ISS² kirjoittavat tallentavat lokitiedon ELF-muodossa. (Vrt. Hallam-Bake & Behlendorf, s.a.)

¹<http://www.apache.org/>

²<http://www.iis.net/>

```
#Software: Microsoft Internet Information Services 6.0
#Version: 1.0
#Date: 2002-05-02 17:42:15
#Fields: date time c-ip cs-username s-ip s-port cs-method
cs-uri-stem cs-uri-query sc-status cs(User-Agent)
2002-05-02 17:42:15 172.22.255.255 - 172.30.255.255
80 GET /images/picture.jpg - 200
Mozilla/4.0+(compatible;MSIE+5.5;+Windows+2000+Server)
```

Lokitiedon merkitys verkkopalvelujen kehittämisessä ja markkinoinnin näkökulmasta kasvoi merkittävästi 2000-luvulle tultaessa. Tähän vaikutti muun muassa Internetin räjähdysmäinen suosion ja yleistymisen kasvu eri puolilla maailmaa. Croll & Sean (2009) mainitsevat eräänä esimerkkinä Internetin suosion kasvusta 1998 järjestettyjen jalkapallon MM-kisojen verkkosivujen käytön. Tapahtuman verkkosivuilla osoitteessa france98.com vieraili kisojen aikana keskimäärin 180 käyttäjää sekunnissa. Tämänkaltaiset yksittäiset tapahtumat olivat osallaan vaikuttamassa verkkopalvelujen käytön seurannan kiinnostavuuden kasvuun. Siirryttäessä 2000-luvulle Internetissä alettiin vähitellen tehdä enenevässä määrin kauppaa. Edelläkävijäyrityksiä olivat muun muassa Pizza Hut³ ja Amazon⁴. Internetin suosion kasvu ja verkkokaupankäynti johtivat siihen, että kiinnostus markkinointilähtöiseen lokitiedon keräämiseen kasvoi edelleen. Käytön seurannassa saatavalla tiedolla voitiin osoittaa olevan taloudellista merkitystä verkkopalveluja ylläpitäville organisaatioille.

Nykyään Internetissä on saatavilla tuhansia erilaisia työkaluja, joiden avulla on mahdollista seurata ja analysoida verkkopalvelun käyttöä. Viime vuosina merkittävimpänä tekijänä on ollut ilmaisohjelmien määrän kasvu, joka on mahdollistanut verkkopalveluiden käytön seuraamisen yhä useammille tahoille. Etenkin vuonna 2005 Googlen⁵ ostama ja vuoden vaihteen 2005-2006 aikana lanseeraama Analytics-palvelu⁶ toi verkkopalvelujen käytön seurannan tavallisten ja yksityisten verkkopalvelujen ylläpitäjien saataville. Google Analytics -palvelu kasvatti käytön seurannan alla olevien verkkopalvelujen määrän tuhansista satoihin tuhansiin. Käytön seurantatyökaluista merkittävimpiä ovat mainitun Google Analyticsin lisäksi: Crazy Egg⁷, Alexa⁸ ja Compete⁹. (Croll & Sean, 2009, ss. 69-70; Kaushik, 2010, s. 3; Gube, 2009; Henry, 2007; *Log Files*, s.a.)

³<http://www.pizzahut.com/>

⁴<http://www.amazon.com/>

⁵<http://www.google.com/>

⁶<http://analytics.google.com/>

⁷<http://crazyegg.com/>

⁸<http://www.alexa.com/>

⁹<http://compete.com/>

2.2 Käytön seurantatieto

Verkkopalvelun käytön seurantatieto on sitä tietoa, jota saadaan järjestelmästä, joka seuraa verkkopalvelun käyttöä. Kerätty tieto voidaan jakaa määrälliseen (kvantitatiivinen) sekä laadulliseen (kvalitatiivinen) käytön seurantatietoon. Verkkopalveluissa kvantitatiivista tietoa on mahdollista kerätä esimerkiksi sivulatausmääriin tai verkkopalvelun käyttäjämääriin liittyen. Tiedon keräämisen kannalta kvantitatiivisen tiedon kerääminen on helpompaa kuin kvalitatiivisen tiedon. Kvalitatiivisen tiedon kerääminen vaatii usein erillisten keräysohjelmien toteuttamista sekä kerätyn tiedon tulkitsemista. Laadullista tietoa voidaan käytön seurantajärjestelmän avulla kerätä liittyen käyttäjän toimintatapoihin, tunteuksiin ja haluihin. (Vrt. Grönfors, 1982, s. 4)

Kvantitatiivisen tiedon kerääminen on helpompaa kuin kvalitatiivisen tiedon, koska kerättävien elementtien määrittely on yksinkertaisempaa ja keräimen tekninen toteutus on helpompaa. Kvantitatiivisen tiedon keräämiseen ei yleensä tarvita käyttäjän osallistumista vaan käytön seuraaminen voidaan hoitaa automaattisesti. Parhaassa tapauksessa verkkopalvelun käyttäjä ei edes huomaa, että hänen toimintaansa verkkopalvelussa seurataan. Määrällisen tiedon keräämistä voidaan tehdä helpommin yleiskäyttöisiä, koska monet, joskaan eivät kaikki, määrällisesti kerättävät asiat ovat yhtenäisiä yli verkkopalvelujen. Määrällistä tietoa tallentavia yleiskeräimiä kuvataan tarkemmin aliluvussa 2.6.

Kvalitatiivisen tiedon kerääminen on vaikeampaa kuin kvantitatiivisen. Kvalitatiivisen tiedon keräämisessä joudutaan yhdistelemään tietoa ja liittämään tämä tieto loogisella tavalla verkkopalvelun toimintoihin. Kvalitatiivinen tieto voi kuvata, miksi käyttäjä käyttäytyi verkkopalvelussa toimimallaan tavalla ja miltä hänestä tuntui toimintansa aikana. Verkkopalveluissa olevat toiminnot ovat hyvin verkkopalvelukohtaisia, joten niiden yleistäminen yli verkkopalvelujen on haastavaa. Kvalitatiivisen tiedon keräämisessä voidaan usein vaatia, että käyttäjä osallistuu tiedon jakamiseen järjestelmälle. Kvalitatiivisen tiedon kerääminen käyttäjiltä onnistuu käyttäjäkyselyiden avulla. Tässä työssä paneudutaan toteutuksen osalta pääosin kvantitatiivisen seurantatiedon keräämiseen. Verkkopalvelujen käytön seurannassa on kuitenkin tarkoituksen mukaista pyrkiä ymmärtämään mitä kvalitatiivisella tiedolla tarkoitetaan, kun puhutaan verkkopalvelujen käytön seuraamisesta.

Tiedon keräämisen lisäksi verkkopalvelun käytön seurantaan liittyy olennaisesti kerättävän tiedon käsittely, koostaminen ja visualisointi. Käytön seurantatiedon käsittely voidaan toteuttaa keräyksen kohteena olevaan verkkopalveluun tai kerätty tieto voidaan viedä järjestelmän ulkopuolelle käsiteltäväksi. Tiedon käsittelyn ja visualisoinnin tehtävä on kasvattaa tiedon rikkausastetta eli parantaa sen hyödynnettävyyttä. Käytön seurantatiedon visualisointia käsitellään luvussa 4.

2.2.1 Kvantitatiivinen tieto verkkopalveluissa

Verkkopalvelun käytön seuranta suunniteltaessa kvantitatiivinen näkökulma tietoon on luonteva lähtökohta, koska useat verkkopalvelujen käytöstä kerättävät tekijät voidaan esittää perustuen asian lukumäärään. Kvantitatiivinen käytön seurantatieto voidaan luokitella kategorioihin sen pohjalta, miten monipuolista ja kattavaa kerätty tieto on ja toisaalta, miten helposti se on hyödynnettävissä. Yksinkertainen tiedon kolmijaottelu voidaan kuvata siten, että on olemassa

- raakaa käytön seurantatietoa
- koostettua käytön seurantatietoa ja
- rikasta käytön seurantatietoa.

Raakaalle käytön seurantatiedolle on ominaista, että siitä on vaikea hahmottaa verkkopalvelun käyttöä ja sen hyödyntäminen on hankalaa, koska tiedon rakenteisuus on alhainen. Raakaa käytön seurantatietoa ei ole koostettu millään tavalla. Raakaasta käytön seurantatiedosta voidaan mainita esimerkkinä HTTP-palvelimen (HTTP-server) keräämä lokitieto. HTTP-palvelimen keräämä lokitieto koostuu levyllä olevista tekstitiedostoista, joissa tieto on tallennettu riveittäin aikajärjestykseen.

Koostettu käytön seurantatieto on esimerkiksi raakaasta käytön seurantatiedosta loogiseksi kokonaisuuksiksi yhdisteltyä tietoa. Tällainen looginen kokonaisuus on päivittäisten vierailijoiden määrä tietyllä verkkopalvelun sivulla. Koostetun käytön seurantatiedon avulla on mahdollista havainnoida miten verkkopalvelua käytetään. Esimerkki koostettua ja rikastettua käytön seurantatietoa tarjoavasta käytön seurantajärjestelmästä on Googlen Analytics -palvelu¹⁰.

Rikas käytön seurantatieto on verkkopalvelun käyttöä kokonaisvaltaisesti kuvaavaa tietoa. Rikkaassa käytön seurantatiedossa yhdistyvät verkkopalvelun ominaisuudet, verkkopalvelua käyttävät käyttäjät ja verkkopalvelussa tehdyt toimet. Rikasta seurantatietoa keräävä järjestelmä vaatii räätälöintiä ja tapauskohtaista määrittelyä, jotta seurantajärjestelmä oppii tuntemaan seurattavana olevan verkkopalvelun, koska rikasta käytön seurantatietoa keräävän järjestelmän tulee tuntea seurattavana olevan verkkopalvelun osat ja niiden toiminnallisuudet osatakseen kerätä niihin liittyvää tietoa. Haluttaessa kerätä verkkopalvelun käytöstä rikasta käytön seurantatietoa täytyy verkkopalveluun toteuttaa sisäinen käytön seurantajärjestelmä, jolla on pääsy kaikkiin verkkopalvelun osiin. Verkkopalvelun sisäisen käytön seurantajärjestelmän toteuttamista kuvataan luvussa 3. (Vrt. Valtionvarainministeriö, 2007, s. 82)

Huomattavaa on, että esitetty tiedon laadun kolmijako raakaan käytön seurantatietoon, koostettuun käytön seurantatietoon ja rikkaaseen käytön seurantatietoon

¹⁰<http://analytics.google.com/>

ei ole suoraan verrattavissa tiedon hallinnassa määriteltyihin tiedon laatukäsitteisiin (tieto, informaatio ja tietämys). Käytön seurannasta kerätyssä tiedossa voidaan kuitenkin tunnistaa samankaltaisia tiedon rikkausasteita ja siten edellä esitetyt tiedon tasot voidaan näihin tiedon laatukäsitteisiin löyhästi rinnastaa. (Vrt. Hieta-Wilkman, s.a.)

Kaushik (2010, ss. 16-18) määrittelee, että verkkopalvelun kvantitatiivinen käytön seuranta ja kvantitatiivinen käytön seurantatieto voidaan nähdä käytön seuraamisen versiona 1.0. Kaushikin mukaan tulevaisuudessa verkkopalvelujen käytön seurannassa tulisi siirtyä 2.0-aikakauteen, jossa käytön seuraaminen keskittyisi enemmän kvalitatiivisen tiedon keräämiseen. Tässä työssä toteutetussa käytön seurantajärjestelmässä on keskitytty käytön seurantatiedon keräämiseen etenkin kvantitatiivisesta näkökulmasta, koska tarve kvantitatiiviselle tiedolle oli olemassa ja kvalitatiivista käytön seurantatietoa tuottavan järjestelmän toteutus osoittautui kvantitatiivisen keräimen ohessa liian haastavaksi.

2.2.2 Seurantatiedon merkitys

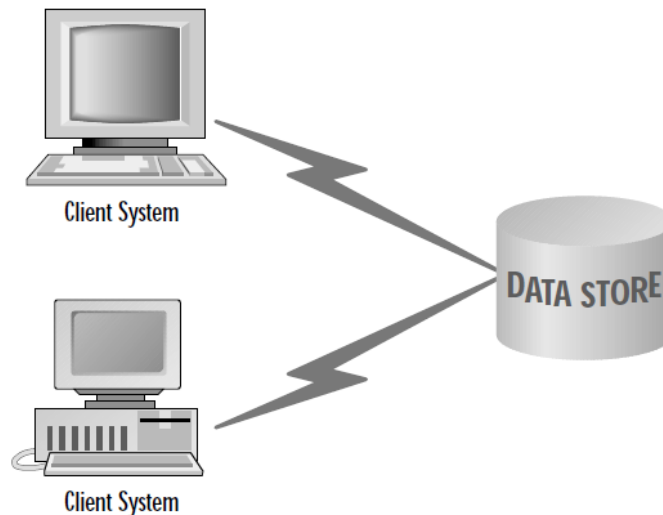
Käytön seurantatiedosta on mahdollista hyötyä erilaisin tavoin ja käytön seurantatietoa voidaan käyttää organisaatioissa apuna monissa tilanteissa. Tässä työssä ei ole tarkoitus kuvata miten käytön seurantatietoa tulisi analysoida ja miten sen perusteella voitaisiin organisaatioissa tehdä strategisia päätöksiä. Käytön seurannan ja siitä saatavan tiedon tärkeys on kuitenkin ymmärtää.

Verkkopalveluja käytetään Internetissä ja prosessina Internet esittäytyy käyttäjille reaaliaikaisena. Internetin reaaliaikaisuus tarkoittaa, että verkkopalveluja ei voida seurata ja tutkia perinteisillä seurantamekanismeilla, jotka tarkastelevat mennyttä aikaa. Jos verkkopalveluun esimerkiksi toteutetaan uusi toiminnallisuus tai muutetaan olemassa olevaa toimintoa, näkyy tämä muutos verkkopalvelussa käyttäjille heti. Tämä tarkoittaa, että verkkopalvelun käytöstä tulee saada reaaliaikaista tietoa. Ilman kattavaa ja helposti saavutettavaa verkkopalvelun käytön seurantatietoa ei organisaatioissa pystytä puuttumaan prosessien ongelmakohtiin ajoissa. Verkkopalvelun käytön seurantatieto on tärkeää verkkopalvelua ylläpitävälle organisaatiolle. (Vrt. Croll & Sean, 2009, ss. 3-5)

2.3 Asiakas-palvelin -malli

Asiakas-palvelin -malli (client-server model) on yleinen tieto- ja viestintäteknologian alalla (Information and Communication Technology, ICT) hyödynnetty malli. Asiakas-palvelin -mallissa käyttäjät selaimineen (browser) tai muine yhteysohjelmineen toimivat asiakkaina. Asiakasohjelmat pyytävät verkossa olevilta palvelimilta käyttäjän etsimää sisältöä. Useimmat Internet-palvelut kuten sähköposti (email),

FTP (File Transfer Protocol) ja WWW-sivut (web pages) toimivat asiakas-palvelin-periaatteen mukaisesti. Verkkopalvelujen eli WWW-liikenteen (World Wide Web) tapauksessa asiakkaana toimii usein selainohjelmisto, kuten Firefox¹¹, ja palvelimena HTTP-palvelin, kuten Apache¹². Kuvassa 2.1 esitellään asiakas-palvelin -mallin yleinen toimintaperiaate. Erilaista verkkohierarkiaa käytetään esimerkiksi toisen sukupolven tai sitä uudemmissa vertaisverkoissa (Peer-to-Peer, P2P), joissa yhteydet luodaan suoraan asiakasohjelmien välille (Bustos-Jimenez et al., 2005, s. 3; Oram, 2001, s. 16).



Kuva 2.1: Internetissä asiakkaita (client system) edustavat selainohjelmistot. Tietovarastona (data store) toimii HTTP-palvelin. Yhteydet luodaan asiakkaiden ja tietovaraston välille. (*Distributed Application Architecture*, 2007, s. 129)

Asiakas-palvelin -mallin rakenteesta johtuen verkkopalvelun käyttöä voidaan seurata kokonaisvaltaisesti, koska yhteydet luodaan asiakasohjelmien ja palvelimen välille. Verkkopalvelun käyttöä on siksi mahdollista seurata palvelimella perustuen asiakasohjelmien palvelimelle lähettämiin pyyntöihin ja palvelimen asiakasohjelmille palauttamiin vastauksiin. Yhteys selaimen (asiakas) ja verkkopalvelun (palvelin) välille muodostetaan käyttämällä HTTP-protokollaa. (*Distributed Application Architecture*, 2007, ss. 129-137)

2.3.1 Asiakasohjelmat

Asiakasohjelmalla (client application) tarkoitetaan ohjelmaa, joka toimii asiakas-palvelin -mallin mukaisesti asiakkaana jollekin palvelinohjelmalle. Asiakasohjelma ja termi asiakas tulee nähdä erillään termistä käyttäjä. Käyttäjällä tarkoitetaan

¹¹<http://www.mozilla-europe.org/firefox/>

¹²<http://www.apache.org/>

asiakasohjelmaa käyttävää henkilöä. Järjestelmää käyttävä käyttäjä ei ole osa asiakas-palvelin -mallia.

Verkkopalvelujen tapauksessa asiakasohjelmalla toimii yleensä jokin WWW-selain. Yksinkertaisimmillaan selain on tekstipohjainen, terminaalikäyttöön soveltuva ja rikkaimmillaan monipuolinen tiedonhaun apuväline. Markkinoilla on tällä hetkellä useita erilaisia selaimia, jotka on tarkoitettu erilaisiin käyttötarkoituksiin ja jotka toimivat kaikki hieman eri tavoin. Selainten versioiden toiminnallisuuksien välillä on suuria eroja. Tyypillisimpiä pöytäkoneissa käytettäviä selaimia ovat tällä hetkellä Firefox¹³, Internet Explorer¹⁴ ja Opera¹⁵ (*Browser Statistics*, 2009; Paul, 2009). Selaimia on kehitetty erilaisten käyttöympäristöjen näkökulmasta, koska perinteisten työpöytäympäristöjen lisäksi verkkopalveluja käytetään lisääntyvässä määrin mobiilipäätelaitteilla. Mobiiliselaimista suosituimpia ovat Applen kehittämä Safari¹⁶ sekä Opera Mini¹⁷, joka on toiminut alalla jo pitkään. (*Opera's history*, 2009; "Opera passes iPhone to lead mobile-browser market", 2009; *Top 9 Mobile Browsers from Dec 08 to Now 09*, 2009)

Verkkopalvelun kehittäjän kannalta selainten moninaisuus on ongelmallista, koska usein selaimet eivät tue samalla tavalla kaikkia WWW-sivuissa käytettäviä toteutusteknologioita, kuten HTML ja CSS (Cascading Style Sheets). Verkkopalveluissa käytettävät toteutusteknologiat eivät siis ole laiteriippumattomia. Pahimmillaan tämä ongelma esiintyy niin, että verkkopalvelu ei ole saavutettava tietyillä selaimilla tai niiden versioilla. Yleensä yhteensopivuusongelmat näkyvät käyttöliittymätason ongelmina eli verkkopalvelussa olevien elementtien asettelu poikkeaa selainten välillä. Verkkopalvelun seurannan kannalta selainten erilaisuus ei ole niin merkittävä ongelma kuin verkkopalveluja kehittäjän tahon näkökulmasta, koska verkkopalvelujen käytön seuranta perustuu pääasiassa alemmalla tasolla toimivien protokollien seurantaan ja nämä protokollat toimivat selaimista riippumattomasti. (*Web browser HTML support*, 2009; *A TCP/IP Tutorial*, 1991; *HTTP - Hypertext Transfer Protocol*, 2009)

2.3.2 Palvelinohjelmat

Palvelinohjelmat (server applications) ovat tietokoneohjelmia, jotka palvelevat asiakasohjelmia niiltä pyydetyillä resursseilla. Asiakasohjelman näkökulmasta HTTP-palvelin esiintyy "mustana laatikkona" (katso aliluku 2.3). Palvelinohjelma toteuttaa asiakasohjelmalta piilossa siltä pyydetyt tehtävät. Asiakasohjelman ja etenkin käyttäjän näkökulmasta verkkopalvelujen tarjoamat resurssit vaikuttavat hyvin

¹³<http://www.mozilla-europe.org/firefox/>

¹⁴<http://www.microsoft.com/windows/internet-explorer/>

¹⁵<http://www.opera.com/>

¹⁶<http://www.apple.com/safari/>

¹⁷<http://www.opera.com/mini/>

yksinkertaisilta, kuten yksittäisiltä HTML-merkatuilla tiedostoilta. Palvelimella HTML-sivu ei kuitenkaan ole pelkkä yksi tiedosto levyllä, joka palautetaan vasteena asiakasohjelmalle, vaan HTML-sivu rakentuu useiden prosessien tuloksena. Näihin prosesseihin saattaa kuulua tietokantakyselyitä ja loogista prosessointia ohjelmointikielillä.

Yleisin HTTP-palvelin on Apache HTTPd Server¹⁸. Apache on avoimeen lähdekoodiin perustuva HTTP/WWW-palvelinohjelma, jonka pohjalta on rakennettu lisäksi muita samaan teknologiapohjaan perustuvia ratkaisuja kuten Apache Tomcat¹⁹ ja Apache Cocoon²⁰. Kirjoittamisen hetkellä Apache HTTPd Server on yleisin web-palvelin noin 50 % markkinaosuudella; toisella sijalla on Microsoft IIS-HTTP-palvelin²¹ (Internet Information Server), jonka markkinaosuus on noin 35%. (*October 2009 Web Server Survey*, 2009)

2.3.3 Asiakas-palvelin -malli verkkopalveluissa

Verkkopalvelut toimivat asiakas-palvelin -mallin periaatteen mukaisesti. Malli toimii verkkopalveluissa käytännössä niin, että käyttäjä syöttää haluamansa resurssin osoitteen eli URI:n (Uniform Resource Identifier) selaimen. URI on verkkopalvelun yksittäisen resurssin tunniste, jonka avulla haluttuun näkymään voidaan tehdä pyyntö. Yleensä URI:t ovat muotoa ”protokolla://palvelin/osoite”. Protokolla kuvaa yhteystavan ja, mille sovellukselle pyyntö ohjataan palvelimella. URI:n palvelin osa kertoo mihin palvelimeen yhteys halutaan ottaa. Osoitteella kerrotaan mikä resurssi palvelimelta halutaan hakea. (Garshol, 1999; Berners-Lee et al., 2004)

Asiakasohjelma lähettää käyttäjän antaman URI-tunnisteen verkkoon. Verkossa pyyntö ohjataan sitä vastaavan palvelimen käsiteltäväksi. Oikean palvelimen löytäminen onnistuu verkossa olevien nimipalvelimien (Domain Name Service, DNS) avulla. Nimipalvelimet tuntevat URI-tunnisteita vastaavat fyysiset verkko-osoitteet ja osaavat siten ohjata pyynnön verkossa eteenpäin. (*How DNS Works*, 2003; Vixie, 2007; Berners-Lee et al., 2004)

Verkossa kulkeva HTTP-pyyntö (HTTP-request) välitetään aina jollain HTTP-metodilla (HTTP-method). Yleisimmin käytetty metodi on nimeltään GET, jonka avulla pyydetään nimensä mukaisesti jotain verkossa olevaa WWW-sivua tai muuta resurssia. Toinen usein, etenkin lomakkeissa, käytetty metodi on POST, jota käytetään tavallisesti, kun halutaan välittää sivulatauspyynnön yhteydessä käyttäjän verkkosivulle syöttämää tietoa palvelimelle. POST ja GET-metodeja käsitellään tarkemmin aliluvussa 2.4.1.

¹⁸<http://www.apache.org/>

¹⁹<http://tomcat.apache.org/>

²⁰<http://cocoon.apache.org/>

²¹<http://www.iis.net/>

HTTP-pyyntöissä lähetetään osoitteen lisäksi HTTP-otsikkokentät (HTTP-headers). Otsikkokentät sisältävät tietoa asiakasohjelmasta sekä pyyntöön liittyviä lisätietoja. HTTP-otsikkokentät eivät ole pakollisia ja luodakseen HTTP-vastauksen (HTTP-response) palvelin vaatii vain onnistuneen pyyntökutsun. HTTP-otsikkokentät ovat kuitenkin käytön seurannan kannalta merkittäviä, koska ne sisältävät tietoa verkkopalvelun käyttäjästä. (Antill, 2007; Croll & Sean, 2009, ss. 69-70; Berners-Lee et al., 2004)

HTTP-pyyntöön perustuen palvelinohjelmisto suorittaa siltä pyydettyt toiminnot ja palauttaa pyyntöä vastaavan HTTP-vastauksen. Vastaus voi olla esimerkiksi HTML-merkattu (Hypertext Markup Language) WWW-sivu, kuva tai video. Huomioitavaa on, että asiakas-palvelin -mallista johtuen asiakasohjelmisto ei tiedä miten palvelin tuottaa siltä pyydetyn resurssin. Esimerkiksi kun asiakasohjelma pyytää palvelimelta WWW-sivua, on palvelimen mahdollista hyödyntää erilaisia toteutusteknologioita ja tietokantaa vastausta luodessaan. Asiakasohjelma näkee vain lopullisen tuloksen eli esimerkiksi HTML-merkatun WWW-sivun. Tämä mahdollistaa sen, että palvelimella voidaan vapaasti muuttaa resurssien toteutustapaa, kuten ohjelmointikieltä tai tietokannanhallintajärjestelmää niin, että se ei vaikuta asiakasohjelmien toimintaan. Verkkopalvelun käytön seuraamisen kannalta tämä tarkoittaa, että verkkopalveluun on mahdollista toteuttaa monipuoliset käytön seurantatyökalut verkkopalvelun käyttäjän tätä mitenkään havaitsematta. (Garshol, 1999; *Hypertext Transfer Protocol*, 1999, ss. 51-54; Berners-Lee et al., 2004)

2.4 Palvelintason käytön seuranta

Palvelintason käytön seurannalla tarkoitetaan tietoja, joita HTTP-palvelin tallentaa sille tulevista pyynnöistä. Verkkopalvelujen tapauksessa palvelintason käytön seuranta perustuu hyvin pitkälle HTTP-protokollan välittämien pyyntöjen tallentamiseen. HTTP on protokolla, jonka avulla välitetään suurin osa WWW-liikenteestä. Toinen usein käytetty WWW-liikenteen protokolla on HTTPS (Hypertext Transfer Protocol Secure), joka on HTTP:n laajennus. HTTPS mahdollistaa salatun yhteyden luomisen asiakasohjelman ja palvelimen välille. HTTPS:ää käytetään yleensä, kun käyttäjä välittää asiakasohjelmalla palvelimelle arkaluontoista tietoa kuten salasanoja. (*HTTP Over TLS*, 2000)

Palvelimen keräämä lokitieto on hyvin raakamuotoista (katso aliluku 2.2.1). Raakamuotoisuudella tarkoitetaan, että kerättyä tietoa ei ole juurikaan jäsennetty eikä tietoalkioita ole nimetty. Tämä johtuu siitä, että lokitiedon korkeampi rakenteisuus kasvattaisi lokitiedon vievän tilan määrää kovalevyllä. Lokitiedoston koon kasvaminen voi johtaa esimerkiksi siihen, että palvelimella käytettävissä oleva kovalevytila loppuu. Lokitiedon käsittely ja lukeminen on nopeampaa, kun tieto

esitetään mahdollisimman yksinkertaisessa muodossa.

Lokitieto kirjoitetaan HTTP-palvelimissa yleensä erilliseen lokitiedostoon rivi riviltä ja pyyntö pyynnöltä. Kerätyn lokitiedon avulla on mahdollista tutkia verkkopalvelun käyttöä. Palvelimen keräämä lokitieto perustuu kuitenkin HTTP-protokollaan ja sen avulla voidaan käytännössä seurata vain verkkopalveluun tulleiden sivulatauspyyntöjen määriä. Esimerkiksi sivulatauksen yhteys verkkopalvelun käyttäjiin tai toimintoihin ei selviä pelkästään palvelimen lokitiedosta, koska näitä koskevat tiedot eivät välity HTTP-protokollan välityksellä. Kvalitatiivista informaatiota on hyvin vaikeaa muodostaa palvelimen keräämästä lokitiedosta, koska raakamuotoiselle tiedolle on vaikeaa muodostaa verkkopalvelukohtaista merkitystä. (Vrt. *Log Files*, s.a.)

2.4.1 HTTP-metodit

HTTP-metodit (HTTP-method) ovat osa HTTP-protokollaa. HTTP-metodien eli HTTP:n käyttämien yhteydenottotapojen avulla hoidetaan tiedonsiirto asiakasohjelmien ja palvelimien välillä. HTTP-protokollan määrittelemät metodit esitellään taulukossa 2.1. Yleisimmin käytetty HTTP-metodi on GET.

Taulukko 2.1: HTTP-metodit ja niiden käyttötarkoitukset. (*Hypertext Transfer Protocol*, 1999, ss. 51-56; Jacobs, 2004)

Nimi	Kuvaus
POST	Käytetään lomakkeen kenttien tietojen välittämiseen asiakasohjelmasta palvelimelle haluttaessa muuttaa palvelimen tilaa.
GET	Useimmin käytetty HTTP-yhteystapa, jolla pyydetään sivu tai muu haluttu resurssi.
OPTIONS	Käytetään palvelimen tai resurssin ominaisuuksien selvittämiseen.
HEAD	Pyydetään vain otsikkotiedot. Voidaan esimerkiksi käyttää selvittää onko sivu muuttunut.
PUT	Sivun tallentaminen.
DELETE	Sivun poistaminen.
TRACE	Palauttaa pyynnön sellaisenaan. Käytetään vikojen selvittämiseen.
CONNECT	Pyydetään yhteyttä, joka jää päälle.

GET on yksinkertainen metodi, jolla asiakasohjelma voi pyytää palvelimelta yksittäisen sivun tai resurssin. Asiakasohjelma lähettää pyynnön tietoliikenneverkon yli palvelimelle hyödyntäen TCP/IP-yhteyttä (Transmission Control Protocol/Internet Protocol). Palvelin käsittelee asiakkaan lähettämän pyynnön, prosessoi siihen sopivan vastauksen ja palauttaa sen verkon yli takaisin asiakkaalle. Tavallisesti

vastaus on tekstimuotoinen HTML-sivu mutta HTTP-vastaus voi olla sisällöltään esimerkiksi kuva- tai ääniresurssi. (*Hypertext Transfer Protocol*, 1999, s. 39; Jacobs, 2004; Korpela, 2003)

HTTP-metodeista erityisesti GET on määritelty idempotentiksi. Idempotentti tarkoittaa, että samanlaiset ja peräkkäiset GET-pyyntöt tuottavat aina yhtäläisen vastauksen. GET-metodissa pyyntöön liittyvät tiedot koodataan URI-tunnisteseen ja se soveltuu siten käytettäväksi verkkopalvelujen ”turvallisissa” toiminnoissa. Esimerkki GET-metodin välittämistä tiedoista on esitetty taulukossa 2.2.

Taulukko 2.2: Esimerkki GET-metodilla välitetyistä HTTP-otsikkotiedoista.

Kenttä	Arvo
User-Agent	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US) AppleWebKit/532.4 (KHTML, like Gecko) Chrome/4.0.237.0 Safari/532.4
Accept	application/xml,application/xhtml+xml,text/html; q=0.9,text/plain;q=0.8,image/png,*/*;q=0.5
Accept-Language	gzip,deflate
Accept-Encoding	ISO-8859-1,utf-8;q=0.7,*;q=0.7

Toinen usein käytetty HTTP-metodi on POST, jota ei ole määritelty idempotentiksi ja siksi se soveltuu sellaisten pyyntöjen välitystavaksi, joilla on tarkoitus muuttaa verkkopalvelun tilaa. POST-metodin tapauksessa pyyntöön liittyviä tietoja voidaan välittää HTTP:n otsikkokenttien lisäksi HTTP-pyyntön runko-osassa. POST-metodilla voidaan toteuttaa verkkopalvelujen ”ei-turvallisia” toimintoja. (*Hypertext Transfer Protocol*, 1999, s. 50; Jacobs, 2004; Korpela, 2003)

2.4.2 HTTP-vastaukset

HTTP-vastaus (HTTP-response) on asiakasohjelman pyyntöön palvelimen toimesta palautettava vaste, jonka avulla asiakasohjelma pystyy päättelemään onnistuiko lähetetyn pyynnön suoritus. Onnistuneeseen sivupyyntöön palvelin palauttaa esimerkiksi vastauksen: ”HTTP/1.1 200 OK”. Vastauksesta asiakasohjelma pystyy lukemaan käytetyn HTTP-protokollan versionumeron (esimerkissä versionumero on: ”1.1”) ja tilaviesti ”200 OK” kertoo pyynnön onnistuneen.

Tilaviesteistä (status message) 2xx-alkuiset kertovat onnistuneesta pyynnöstä. Näistä tärkeimpiä ovat mainittu ”200 OK” ja ”206 Partial Content”, joka kertoo vastauksen olevan osa pyydettyä tiedostoa (pyydettyessä osaa suuremmasta tiedostosta). 3xx-alkuiset koodit kertovat sisällön löytyvän nykyään muualta kuin sieltä mistä asiakasohjelma niitä pyysi. Koodi ”301 Moved Permanently” kertoo sisällön

siirtyneen pysyvästi ja “302 Moved Temporarily”, että sivu on siirretty väliaikaisesti toiseen paikkaan. 4xx-alkuiset vastauskoodit kertovat asiakasohjelman tekemästä virheestä. Esimerkiksi “404 Not Found” on tuloksena, jos asiakasohjelma hakee sivua, jota palvelimella ei ole ja vastaus “403 Forbidden” merkitsee, että asiakasohjelma pyysi sisältöä, johon sillä ei ole tarvittavia oikeuksia. 5xx-alkuiset koodit ovat palvelimen tekemien virheiden seurausta. Näistä yleisin on 500 “Internal Server Error”, joka kertoo pyynnön epäonnistuneen sitä käsitelleen ohjelman kaaduttua tai palvelimen muun virheellisen toiminnan seurauksena. Kuvassa 2.2 näytetään esimerkki visualisoinnista, jossa esitetään eräässä verkkopalvelussa yleisimmät HTTP-vastauksien virhekoodit.

HTTP-virhekoodit*		Osumia	Prosentti	Kaista
403	Forbidden	198	51.5 %	54.75 Kt
302	Moved temporarily (redirect)	134	34.8 %	33.02 Kt
404	Document Not Found	38	9.8 %	11.11 Kt
400	Bad Request	6	1.5 %	1.77 Kt
301	Moved permanently (redirect)	5	1.3 %	1.54 Kt
401	Unauthorized	3	0.7 %	1.41 Kt

Kuva 2.2: Palvelimen palauttamien HTTP-vastauksien määriä on mahdollista seurata verkkopalveluissa.

Asiakasohjelmat hyödyntävät vastauskoodeja verrattain vähän, vaikka niihin perustuen voitaisiin toteuttaa käyttäjän toimintaa helpottavia ominaisuuksia. Asiakasohjelman olisi esimerkiksi mahdollista sopivan vastauskoodin saadessaan havaita, että pyydetty resurssi on siirretty toiseen osoitteeseen ja päivittää asiakasohjelmaan tallennettu kirjanmerkki automaattisesti resurssin uutta osoitetta vastaavaksi. Epäonnistuneen HTTP-pyyntöön yhteydessä useimmat asiakasohjelmat palauttavat epäselvän virheilmoitusivun, joka sisältö on pahimmillaan pelkkä HTTP-vastauksen tilaviestinumero. (*Hypertext Transfer Protocol*, 1999, s. 40; Croll & Sean, 2009, ss. 69-70)

2.4.3 Apache

Apache²² on HTTP-palvelinohjelmisto, jonka historia ulottuu vuoteen 1995, jolloin yleisin HTTP-palvelinohjelmisto oli NCSA:n (National Center for Supercomputing Applications) kehittämä HTTPd²³. HTTPd kehitystyö kuitenkin pysähtyi vuoden 1994 puolivälissä, kun palvelimen pääkehittäjä erosi NCSA:n palveluksesta. HTTPd oli tällöin suosituin käytetyistä HTTP-palvelinohjelmistoista, joten sen kehitykselle oli tarvetta. Tästä johtuen monet NCSA:sta riippumattomat tahot jatkoivat HTTPd kehittämistä. Kehitystyö ei ollut kuitenkaan koordinoitua. HTTPd kehitystyötä

²²<http://apache.org/>

²³<http://hoohoo.ncsa.illinois.edu/>

jatkaneet tahot kokoontuivat pian yhteen ja lopputuloksena syntyi Apache - Apache Patchy Server sekä Apache Group. Apachen ensimmäinen virallinen versio 0.6.2 perustui hyvin pitkälle HTTPd:n versioon 1.3. ja se julkaistiin huhtikuussa 1995. Apachen suosion kasvu oli räjähdysmäistä ja vuodessa julkaisemisensa jälkeen siitä tuli suosituin HTTP-palvelinohjelmisto. Apachen suosion ja menestyksen taustalla on useita tekijöitä:

- Apache on erittäin nopea, varma ja stabiili palvelinohjelmisto
- Apachen lähdekoodi on avointa. Avoimen lähdekoodin ja lisensiointikäytännön ansiosta Apache on saanut merkittävän jalansijan useissa kaupallisissa palvelinratkaisuissa kuten Red Hat -palvelinsarjan²⁴ tuotteissa
- Apache on saatavilla useimmille alustoille (Windows, *Nix ja OS/2.) ja
- Apache on ilmainen toisin kuin useimmat muut raskaaseen käyttöön suunnitellut HTTP-palvelimet. Yrityskäyttäjille on tarjolla maksullisia tukipalveluita ja ratkaisuja. (*How Apache Came to Be*, 2009; *NCSA HTTPd*, 1998; *October 2009 Web Server Survey*, 2009)

Apache kerää lokitietoa verkkopalvelun käytöstä sivulatauksittain asetustiedostossa määriteltäviin sijainteihin levyllä. Apachen keräämä lokitieto sisältää kaiken oleellisen tiedon liittyen WWW-sivulataustapahtumiin HTTP-protokollan tasolla. Apache kirjoittaa lokimerkinnot ELF-muodossa. Alla esimerkki Apachen lokitiedoston keräämästä lokitietomerkinnästä:

```
130.230.8.13 -- [28/Sep/2009:10:41:13 +0300]
"GET /wiki/index.php/Statster_v.3.1 HTTP/1.1"
200 8525 "http://teelmo.info/wiki/index.php?title=Statster_v.3.1"
"Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.1.3)
Gecko/20090824 Firefox/3.5.3 (.NET CLR 3.5.30729)"
```

Taulukossa 2.3 lokitieto on jäsennettynä, että jokaiselle tietoalkiolle on annettu nimi, jotka on eroteltu toisistaan omille riveilleen. Kokenut verkkopalvelun ylläpitäjä pystyy tulkitsemaan Apachen lokitietoa sellaisenaan. Lokitieto sisältää tietoa HTTP-pyyntöön lähettäneestä asiakasohjelmasta ja HTTP-pyyntöön liittyvistä lisätiedoista.

²⁴<http://www.redhat.com/>

Taulukko 2.3: Jäsennetty Apachen keräämä lokitieto.

Kenttä	Arvo
Käyttäjän IP-osoite	130.230.8.13
Aikaleima	28/Sep/2009:10:41:13
Aikaero	+3h
HTTP-yhteystapa	GET
Sivun URL	/wiki/index.php/Statster_v.3.1
HTTP-versio	1.1
HTTP-vastaus	200 (OK)
Sivun URI	http://teelmo.info/wiki/index.php?title=Statster_v.3.1
Käyttöjärjestelmä	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.1.3)
Asiakasohjelma	Gecko/20090824 Firefox/3.5.3 (.NET CLR 3.5.30729)

Taulukon 2.3 kaltaisesta lokitietomerkinästä pystytään tulkitsemaan, että käyttäjä “130.230.8.13” on pyytänyt palvelimelta wiki-sivua nimeltä “Statster_v.3.1” ja pyyntö on onnistunut. Käyttäjän asiakasohjelmassa on ollut Firefox²⁵ ja käyttöjärjestelmänä on toiminut Windows²⁶. Lokitiedossa on lisäksi aikaleimatieto pyynnön tapahtumisajasta. (Vrt. Croll & Sean, 2009, ss. 114-115)

2.5 Asiakasohjelman käytön seuranta

Verkkopalvelun käyttöä voidaan seurata asiakasohjelmassa ajettavilla komentosarjakielillä. Yleinen Internetissä käytetty komentosarjakieli on JavaScript. Asiakasohjelmiin kytkeytyvä käytön seuranta mahdollistaa tiettyjen tietojen keräämisen, joiden kerääminen ei esimerkiksi palvelintasolla toimivalla seuraimella ole mahdollista. Toisaalta kun käytön seurantatiedon keräys riippuu käyttäjän asiakasohjelmasta on mahdollista, että kaikki kerättävänä oleva tieto ei tule tallennetuksi. Tähän vaikuttavat esimerkiksi käyttäjän asiakasohjelmaan tekemät asetukset tai asiakasohjelman tekniset rajoitteet.

Asiakasohjelmassa käytön seuraaminen tapahtuu siten, että asiakasohjelma suorittaa palvelimella olevan käytön seurantatiedon keräävän komentosarjan. Tämän jälkeen asiakasohjelma lähettää tiedot verkon yli käytön seurantatietoa tallennettavalle palvelimelle. Tietojen kerääminen ja palvelimelle lähettäminen on suoritettava verkkosivun muiden osien lataamisen jälkeen, jotta käytön seuraamisella olisi mahdollisimman pieni vaikutus verkkopalvelun normaaliin käyttöön. Asiakasohjelmaan kytkeytyvän käytön seurannan etuna on, että seurantatietoa saadaan kerättyä

²⁵<http://www.mozilla-europe.org/firefox/>

²⁶<http://windows.microsoft.com/>

liittyen käyttäjän tekniseen käyttöympäristöön, johon kuuluvat käyttäjän resoluutio, selaimen tiedot, käyttäjän näytön väripaletti ja käyttöjärjestelmätiedot.

Asiakasohjelmissa toteutettava käytön seuranta, joka perustuu komentosarjakieliin on “haavoittuva”, koska käyttäjän on mahdollista kytkeä se halutessaan pois päältä. Kaikissa nykyaikaisissa asiakasohjelmissa on esimerkiksi mahdollista estää verkkopalvelussa olevien JavaScript-komentosarjojen suoritus. Jos käytön seurannan toteuttava komentosarja ei tule ajetuksi tarkoittaa tämä, että käyttäjän toimia verkkopalvelussa ei pystytä tallentamaan ja on mahdollista, että käyttäjän asiakasohjelma ei tue valittua käytön seurantateknologiaa. Tässä tapauksessa kyseessä olevat päätelaitteet jäävät seurantatiedon ulkopuolelle.

2.6 Käytön seurantajärjestelmä

Käytön seurantajärjestelmällä tarkoitetaan järjestelmää, joka toteuttaa verkkopalvelussa käytön seurantaa joko kokonaisvaltaisesti tai osittain esimerkiksi vain tiedon koostamisen osalta. Tämänkaltaisia järjestelmiä on tällä hetkellä markkinoilla useita. Järjestelmien monipuolisuus, käyttöönoton vaikeus ja toimintaperiaatteet vaihtelevat järjestelmien kesken suuresti. Yhteistä järjestelmille on pyrkimys koostaa verkkopalvelun käytöstä tietoa. Koostetun tiedon tarkoituksena on tehdä verkkopalvelun käyttö näkyväksi. Järjestelmien suuri määrä johtuu siitä, että yksinkertaisen käytön seurantajärjestelmän toteuttaminen vaatii vain vähän teknistä osaamista, koska käytön seurantatietona voidaan hyödyntää HTTP-palvelimen keräämää lokitietoa. Käytön seurantajärjestelmän tehtäväksi jää tällöin vain tiedon koostaminen haluttuun muotoon. Syynä siihen, miksi järjestelmiä on niin useita voidaan nähdä, että käytön seurantatiedon tarve vaihtelee hyvin paljon verkkopalvelujen välillä.

Tietyllä tasolla valmiit käytön seurantajärjestelmät tarjoavat hyvin pienellä teknisellä panostuksella kattavia tilastoja verkkopalvelun käytöstä. Pienellä panostuksella siitä näkökulmasta, että vaihtoehtona on käytännössä toteuttaa käytön seurantaan liittyvä toiminnallisuus täysin itse. Valmiiden järjestelmien ongelmana on kuitenkin se, että ne seuraavat pääasiassa pelkkiä sivulataustapahtumia ja tekevät sen hyvin yleisellä tasolla. Verkkopalvelut sisältävät usein ominaisuuksia, joiden seuraaminen ei ole mahdollista, jos käytön seurain ei niitä erikseen tunnista. Esimerkiksi sivulataustapahtuman liittäminen verkkopalvelun käyttäjätiliin ei ole mahdollista olemassa olevilla käytön seurantajärjestelmillä. Järjestelmän toimintojen suosio ja käyttö ovat yleensä näiden järjestelmien seurannan ulottumattomissa, koska ne eivät tunne verkkopalvelukohtaisia toimintoja eivätkä siten pysty rekisteröimään niitä. Mainittujen tekijöiden vuoksi kattavan ja kokonaisvaltaisen käytön seurantatiedon keräämiseksi on verkkopalveluun toteutettava oma seuranta-

järjestelmä, jolla on pääsy verkkopalvelun kaikkiin osiin. Esimerkiksi verkkopalvelun tietokantaan tallennettavat tiedot ovat usein verkkopalvelua ulkopuolelta seuraavien järjestelmien saavuttamattomissa. Voidaan yleistää niin, että valmiit käytön seurantajärjestelmät keräävät ja koostavat kattavasti sivulatauksiin liittyvää tietoa, mutta verkkopalvelukohtainen tieto eli tieto, mitä käyttäjä todellisuudessa verkkopalvelussa teki jää niiltä piiloon.

Käytön seurantajärjestelmät toimivat useimmiten seuraavaksi kuvattavilla kahdella toisistaan poikkeavalla tavalla. Ensinnäkin ne voivat toimia kuten edellä kuvattu eli seuraamalla HTTP-palvelimen keräämää lokitietoa. Tässä tapauksessa käytön seurantajärjestelmän toiminnallisuus keskittyy lähinnä siihen miten HTTP-palvelimen keräämä tieto voitaisiin esittää mahdollisimman hyödyllisessä muodossa. Tämän kaltainen käytön seurantajärjestelmä on sijoitettu tavanomaisesti samalle palvelimelle kuin missä itse verkkopalvelu toimii. HTTP-palvelimen lokitiedostoa seuraavien järjestelmien käyttöönotto vaatii tiettyä tietoteknistä osaamista, koska järjestelmien asentamisen yhteydessä käsitellään tekstipohjaisia asetustiedostoja. Asetustiedostojen muokkaaminen ei välttämättä ole luontevaa graafisiin asennusohjelmiin ja ympäristöihin tottuneille käyttäjille. Seurantajärjestelmää käyttöönottavalla organisaatiolla tulee olla hallussaan sellainen palvelinympäristö, johon käytön seurantajärjestelmä on mahdollista asentaa. Asennusoikeuksien lisäksi käytön seurantajärjestelmällä tulee olla pääsyoikeus HTTP-palvelimen keräämiin lokitietoihin. Käyttöoikeuksien riittämättömyys voi tulla esiin toimittaessa ulkopuolisilta tahoilta hankituissa ympäristöissä kuten webhotelleissa (web hosting) tai virtuaalipalvelimillä (virtual server). AWStats on esimerkki HTTP-palvelimen lokitietoa koostavasta seurantajärjestelmästä.

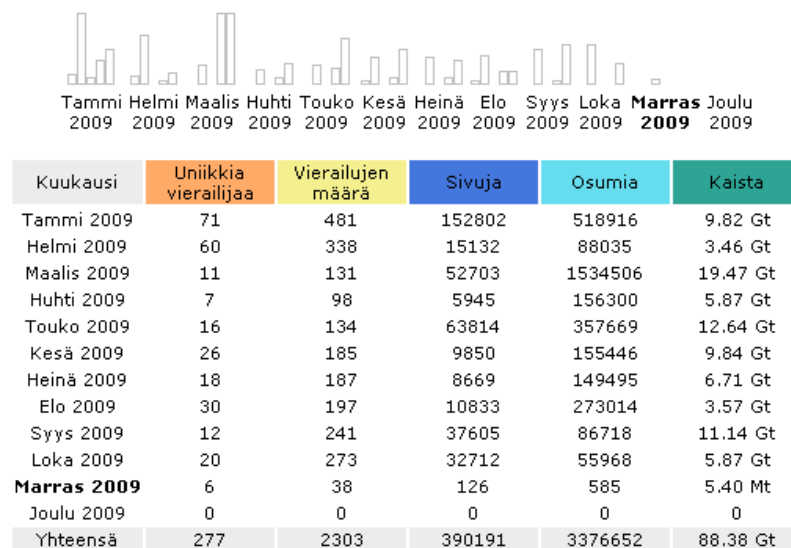
Toinen yleisesti verkkopalvelujen käytön seurantajärjestelmien hyödyntämä keino seurata verkkopalvelun käyttöä on kerätä tieto sivulatauksesta käyttäjän asiakasohjelmassa (katso aliluku 2.5). Verkkopalvelun seuranta tapahtuu tässä tapauksessa niin, että verkkopalvelun kehittäjä lisää verkkopalveluunsa koodilohkon, joka aktivoituu jokaisen sivulatauksen yhteydessä. Koodilohko kerää sivulataukseen liittyvät tiedot ja lähettää ne, yleensä salattuna, eteenpäin käytön seurantapalvelimelle. Käytön seurantatietojen säilyttäminen ja käsittely tapahtuu tässä tapauksessa käytön seurantapalvelua tarjoavan tahon palvelimella. Positiivisena puolena käytön seurantatietoon liittyvä laskenta ja seurantaan liittyvät ylläpidolliset tekijät ulkoistuvat. Negatiivisena puolena voidaan käytön seurantatiedon antaminen kolmannen osapuolen (1. käyttäjä, 2. verkkopalvelu ja 3. käytön seurantajärjestelmä) haltuun. Tämä muodostuu ongelmaksi esimerkiksi silloin jos verkkopalvelun käyttäjät eivät luota tähän kolmanteen osapuoleen. Luottamuksen puute saattaa johtaa siihen, että verkkopalvelun käyttäjät rajoittavat toimintaansa verkkopalvelussa, koska eivät halua, että tämä kolmas osapuoli saa haltuunsa tietoa heidän

verkkopalvelun käytöstään. Verkkopalveluun liitettävän ja asiakasohjelmassa suoritettavan käytön seuraimen toimintaan asettaminen ei vaadi verkkopalvelua ylläpitävältä organisaatiolta erityistä osaamista tai oikeuksia palvelimella. Seurantajärjestelmän vaatima koodilohko sijoitetaan muun olemassa olevan koodin joukkoon, johon verkkopalvelun kehittäjällä on muutenkin oikeus.

Edellä kuvattujen toimintatapojen lisäksi verkkopalvelun kehittäjälle on mahdollista toteuttaa verkkopalveluun erillinen käytön seurantajärjestelmä, jonka avulla seuranta toteutetaan. Käytön seurantajärjestelmän toteuttamista verkkopalvelun yhtehteen kuvataan tarkemmin luvussa 3.

2.6.1 AWStats

AWStats²⁷ on avoimeen lähdekoodiin perustuva verkkosivujen käytön seurantajärjestelmä, joka on kirjoitettu PERL-ohjelmointikielellä (Practical Extraction and Report Language). AWStats jäsentää ja koostaa palvelimen, esimerkiksi HTTP-palvelimen, lokitiedostoja ja tuottaa niistä HTML-muotoisia raportteja (kuva 2.3). AWStatsin käyttö on käyttäjälle näkymätöntä eikä käyttäjällä ole mahdollista vaikuttaa sen toimintaan. Tämä siis erona esimerkiksi Google Analyticsiin, jossa käyttäjä näkee ja pystyy estämään yhteyden Googlen palvelimelle sivulataustapah-tuman yhteydessä. AWStats on yksi monipuolisimmista HTTP-palvelimen lokitieto- ja jäsentävistä käytön seurantajärjestelmistä. (*Log analyzers Comparisons*, 2009)



Kuva 2.3: AWStatsin tuottama visualisointi verkkopalvelun käytöstä kuukausittain.

AWStatsilla on mahdollista seurata verkkopalvelujen käytön lisäksi muiden Internet-palveluiden käyttöä. AWStats osaa siis jäsentää muidenkin kuin HTTP-

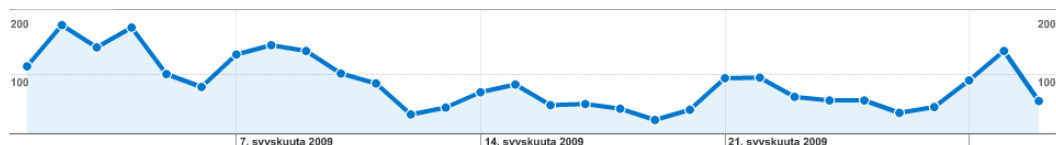
²⁷<http://awstats.sourceforge.net/>

palvelimien luomaa lokitietoa. Esimerkiksi FTP-palvelimen (File Transfer Protocol server) liikennettä, sähköpostia ja virtaavaa mediaa (streaming media) kuten Internet-radion (web radio) käyttöä on mahdollista seurata AWStatstin avulla.

2.6.2 Google Analytics

Google Analytics²⁸ on Googlen²⁹ omistama verkkosivujen kävijäseurantaan ja tiedon visualisointiin erikoistunut verkkopalvelu. Google hankki ohjelmiston Urchin Software Corporationilta vuonna 2005. Analytics-palvelu julkaistiin koekäyttöön vuoden 2005 lopussa ja täysimittaisesti vuoden 2006 aikana.

Palvelua käyttäekseen verkkopalvelun ylläpitäjän täytyy luoda tunnus Google Analytics -palveluun. Palvelussa on tämän jälkeen mahdollista luoda verkkopalvelukohtaisia profiileja. Profilit erotetaan toisistaan yksikäsitteisellä avaintunnisteella. Analytics-palvelu toimii siten, että verkkopalvelun ylläpitäjän on mahdollista ladata itselleen valmis JavaScript-koodilohko, joka on yksilöity edellä mainitulla profiilikohteisella avaimella. Koodilohko sijoitetaan verkkopalvelun lähdekoodiin, josta asiakasohjelma suorittaa sen sivulataustapahtuman yhteydessä ja lähettää keräämänsä sivulataukseen liittyvät tiedot Analytics-palveluun. Koodilohko voidaan halutessa sijoittaa vain osaan verkkopalvelun sivuista, jos seuranta ei haluta tehdä verkkopalvelun kaikista osista. Tällä tavalla on mahdollista jättää esimerkiksi verkkopalvelun ylläpitoalue käytön seurannan ulkopuolelle. Analytics-palvelu luo kerätyistä tiedoista erilaisia visualisointeja, jotka ovat saatavilla kirjautumalla Analytics-palveluun. Kuvassa 2.4 nähdään yksinkertainen verkkopalvelun sivulatausmääriä kuvaava visualisointi Analytics-palvelussa.



Kuva 2.4: Esimerkki sivulatausten määrää kuukauden ajalta kuvaavasta viivadiagrammista Google Analytics -palvelussa.

Google Analyticsin seurannan toteuttava koodilohko on toteutettu JavaScript-ohjelmointikielellä. Tämä tarkoittaa, että Analytics-seurantajärjestelmä ei pysty keräämään tietoa niistä käyttäjistä, joiden selaimet eivät tue JavaScriptiä, jotka ovat kytkeneet JavaScriptin pois käytöstä tai jotka ovat muulla tavoin estäneet Google Analytics seurannan. Esimerkiksi useat mobiiliselaimet eivät tue JavaScriptiä. JavaScript-tuki puuttuu lisäksi tekstipohjaisista selaimista. Käytön seuranta-tietojen ja niiden visualisointien sijoittaminen Google Analytics-palvelun käyttämi-

²⁸<http://analytics.google.com/>

²⁹<http://www.google.com/>

nen voi olla lisäksi ongelmallista, jos palvelun keräämiä tietoja halutaan käsitellä Analytics-palvelun ulkopuolella. (*Google Analytics*, 2010)

Google Analytics on verkkopalveluissa käytettävistä käytön seurantapalveluista suosituin (*The Biggest Google Analytics Sites*, 2009). Syynä tähän ovat palvelun yksinkertainen käyttöönotto, ilmaisuus sekä kattavat, jopa ammattilaiskäyttöön soveltuvat käytön seurannan visualisointityökalut.

2.7 Sovellustason käytön seuranta

Verkkopalvelun sovellustason käytön seurannalla tarkoitetaan sellaista toiminnallisuutta, joka kerää verkkopalvelun käytöstä tietoa verkkopalvelun ”sisältä käsin”. Sovellustason käytön seuranta toteutetaan osaksi verkkopalvelun muutakin toiminnallisuutta ja sillä on siten mahdollisuus päästä käsiksi laajemmin verkkopalvelun tietosisältöihin kuin esimerkiksi HTTP-palvelimella tai asiakasohjelmassa toimivalla käytön seuraimella. Termistönä on hyvä huomioida, että sovellustason käytön seuranta tapahtuu palvelimella ja usein fyysisesti samalla koneella, jolla HTTP-palvelinohjelma toimii ja kerää tietoa. Sovellustason käytön seuranta toteutetaan kuitenkin sovellusten sisään, jotka toteuttavat HTTP-palvelimelta pyydettyt resurssit ja niiden on siten mahdollista kerätä tietoa laajemmin.

Sovellustason käytön seurantajärjestelmälle on ominaista ainutlaatuisuus. Yhteen verkkopalveluun toteutettua käytön seurantajärjestelmää ei voida sellaisenaan hyödyntää missään muussa verkkopalvelussa, koska verkkopalvelut ovat toteutettu eri tavoin ja ne sisältävät erilaisia toimintoja. Käyttäjien toimien seuraamiseen toteutettu käytön seurantajärjestelmä ei siten voi toimia kuin siinä verkkopalvelussa, johon se on kehitetty. Hyviä käytäntöjä ja joitain käytön seurantajärjestelmän osia voidaan hyödyntää tai jopa siirtää verkkopalvelujen seurantajärjestelmien kesken, mutta sovellustasolla toimiva käytön seurannan toteuttava kokonaisuus on aina jollain tavalla räätälöitävä kyseessä olevaan verkkopalveluun sopivaksi.

2.7.1 Tietokannan rooli sovellustason käytön seurannassa

Verkkopalvelun käytön seurantajärjestelmän tulee pystyä tallentamaan keräämänsä tiedot pysyvään tallennuspaikkaan. Usein kerätty käytön seurantatieto tallennetaan tietokantaan (database). Tietotekniikan alalla tietokanta on termi, jota käytetään haluttaessa viitata paikkaan, johon tietoa voidaan varastoida. Tietokanta voi olla periaatteessa missä tahansa muodossa ja esimerkiksi kalenteri on eräänlainen tietokanta eli tietovarasto. Tietokannalle on ominaista, että tietokannassa olevilla tiedoilla voi olla yhteyksiä keskenään.

Verkkopalvelujen kohdalla tietokantana toimii usein SQL-pohjainen (Structured Query Language) tietokannan hallintajärjestelmä. SQL-pohtaiset tietokannat

ovat tyypiltään relaatiotietokantoja ja niihin voidaan tehdä relaatioalgebraan perustuvia hakuja. SQL on suosittu verkkopalveluissa, koska sen käyttö on sisäänrakennettu useimpiin verkossa hyödynnettyihin toteutusteknologioihin ja niissä käytettyihin ohjelmointikieliin. SQL-pohjaisia tietokannan hallintajärjestelmiä ovat muun muassa MySQL³⁰ ja SQLite³¹, joista MySQL on tällä hetkellä verkkopalveluissa käytetyistä avoimeen lähdekoodiin perustuvista tietokannan hallintajärjestelmistä suosituin. Tietokanta voi perustua tietokantaan tallennettaviin objekteihin. Rakenteeltaan tietokanta voi olla hajautettu tai keskitetty. (Vrt. Hiltunen, 1999, ss. 16-29; *Market Share*, s.a.)

Tietokanta toimii siis käytön seurantatiedon tallennuspaikkana. Tiedot tallennetaan tietokantaan rakenteisessa muodossa, joka tarkoittaa, että tallennetulla tiedolla on määrittely ja tieto on nimettyä. Tiedon rakenteisuus helpottaa käytön seurantatiedon käyttämistä ja jalostamista.

2.7.2 Seurattavien toimintojen tunnistaminen sovellustasolla

Verkkopalvelut sisältävät erilaisia toimintoja, joiden avulla verkkopalvelua käytetään. Toimintoja käytetään tietyn verkkopalvelun sivun eli näkymän kautta. Näkymän, jolla toiminto sijaitsee, lataaminen ei kuitenkaan tarkoita, että näkymässä oleva toiminto suoritettaisiin. Jos esimerkiksi verkkopalvelussa on näkymä “/rekisteröityminen/”, joka toteuttaa toiminnallisuuden ja jonka avulla verkkopalveluun voidaan luoda uusi käyttäjätunnus, ei näkymän lataustietojen perusteella voida päätellä verkkopalveluun rekisteröityneiden käyttäjien määrää. Usein toiminnot suoritetaan verkkopalvelussa lomakkeiden avulla. Käyttäjä täyttää lomakkeessa kysyttävät tiedot ja lähettää ne palvelimelle. Tällöin toimintoja voidaan seurata tallentamalla sivulataustapahtuman yhteyteen tieto siitä, onko sivulla lähetetty lomake. Toimintojen seuraamiseksi täytyy seurantajärjestelmän siis kyetä tunnistamaan verkkopalvelun käytöstä sivulataustiedon lisäksi muita tekijöitä. Käytön seurannassa toiminnon tunnistaminen voidaan tehdä joko tiedon keräysvaiheessa tai vaihtoehtoisesti käytön seurantatiedon käsittelyn yhteydessä. Jos tunnistaminen tehdään jo tiedon keräysvaiheessa, voidaan käytön seurantatietoon tallentaa suoraan esimerkiksi toiminnon tunniste.

2.7.3 Sovelluksen käyttäjätunnuksen liittäminen asiakasohjelmaan

Sovellustasolla tapahtuvan käytön seurannan etuna on, että käytön seurantatiedon sisältämään sivulatausten sarjaan voidaan yhdistää tieto verkkopalvelussa toimivista käyttäjätunnuksista. Toisin sanoen verkkopalvelun sivun ladannut asiakasohjelma

³⁰<http://www.mysql.com/>

³¹<http://sqlite.org/>

on mahdollista yhdistää verkkopalveluun rekisteröityneeseen henkilöön. Liittäminen voidaan tehdä esimerkiksi evästeiden (cookies) avulla. Evästeet voidaan tallentaa käyttäjän asiakasohjelmaan ja niiden avulla on mahdollista yhdistää asiakasohjelma palvelun käyttäjätunnukseen käyttäjän kirjautuessa. Kun verkkopalvelun käyttöä seurataan esimerkiksi HTTP-palvelimella tai asiakasohjelmassa ajettavilla komentosarjakielillä, on käyttäjä mahdollista yksilöidä vain laitteen teknisten ominaisuuksien perusteella. Esimerkki käyttäjän teknisestä ominaisuudesta on käyttäjän asiakasohjelman käyttämä IP-osoite. Verkkopalvelua selaavien henkilöiden teknisiä tietoja ja verkkopalvelun käyttäjätunnuksia yhdistämällä on mahdollista tutkia, mistä fyysisistä sijainneista käyttäjät käyttävät verkkopalvelua. Toisaalta käyttäjän IP-osoitetta ja käyttäjätunnusta tutkimalla voidaan selvittää, mikä on verkkopalvelussa ominainen polku, joka johtaa kirjautumiseen. Tämä onnistuu tarkastelemalla kirjautuneen käyttäjän IP-osoitteen toimia verkkopalvelussa ennen käyttäjän kirjautumista.

2.7.4 Käytön seurannan tehokkuuskysymykset

Käytön seurantajärjestelmän tehokkuus vaikuttaa siihen miten verkkopalvelun käyttäjä kokee käytön seurannan. Seurantajärjestelmän tulisi olla käyttäjälle näkymätön prosessi, joka ei vaikuta verkkopalvelun normaaliin käyttöön eli tallennusprosessin tulisi olla mahdollisimman kevyt. Seurantatiedon tallentaminen ei siis saa vaikuttaa merkittäväällä tavalla esimerkiksi verkkopalvelun latausaikoihin. Hyvä periaate on tehdä käytön seurantatiedon tallennus vasta, kun sivun muu sisältö on ladattu ja saatu esitettyä käyttäjälle. Kun käytön seurantatieto lähetetään toiselle palvelimelle, kestää tämä lähetys TCP:n kolmikättelyn johdosta aina suhteellisen pitkään. Tämän yhteyden ottamisen ja tiedon lähettämisen aikana sivun muun sisällön lataaminen ei välttämättä etene eikä käyttäjä siten pääse käyttämään pyytämäänsä verkkosivua. Pahimmassa tapauksessa yhteyden luonti saattaa epäonnistua ja siitä seuraava aikakatkaus voi kestää useita sekunteja. (Vrt. Kozierok, 2005, ss. 752-765)

Käytön seurantatiedon tallentaminen voi joissain verkkopalveluissa ja tapauksissa vaatia paljon laskentaa kuten kaikkien tapahtumaan liittyvien tietojen kerääminen. Tämän prosessoinnin määrä tulisi kuitenkin pyrkiä minimoimaan mahdollisuuksien mukaan. Jos tiedon prosessoinnin määrää ei voida vähentää, tulisi sen suoritus siirtää pois sivulataustapahtuman yhteydestä esimerkiksi eräajettavaksi erillisenä ajankohtana.

2.8 Käytön seurannan vaikutukset verkkopalvelun käyttäjään

Käytön seurannalla on aina vaikutus verkkopalvelun käyttäjään joko suoraan tai välillisesti. Suoraan verkkopalvelun seuranta voi vaikuttaa esimerkiksi pidentyvinä

sivulatausaikoina. Välillisesti käytön seuraaminen voi vaikuttaa käyttäjän yksityisyyteen, kun käyttäjästä kerätään häneen yhdistettäviä tietoja. Yksilötason seuranta saattaa asettaa verkkopalvelun outoon asemaan käyttäjän näkökulmasta, jos käyttäjä tuntee olevansa valvovan silmän alla. Yksilötason seuraamista kannattaa yleensä välttää ja pyrkiä keskittymään käytön seuraamisessa ja etenkin sen visualisoinnissa suurempien kokonaisuuksien käsittelyyn. Yksityisyyteen liittyvät ongelmat vähenevät, kun verkkopalvelun käyttöä visualisoidaan, jolloin yksittäisen käyttäjän toimet eivät ole yksilöitävissä,

Kokonaisuuksien kuvaaminen käytön seurantatiedon avulla onkin järkevämpää jos sen perusteella halutaan tehdä johtopäätöksiä. On perustellumpaa tehdä johtopäätöksiä kokonaisuuksiin perustuen kuin yksittäisiin käyttäjiin. Huomioitava on yksittäisten seurantatietoalkioiden sisältö ja se kenellä niihin on pääsy. Jos kerättävä tieto on niin tarkkaa, että sen avulla voidaan yksilöidä käyttäjä, täytyy tieto olla vain sellaisten henkilöiden saavutettavissa, jolle se kuuluu.

Hyvä käytäntö on kertoa verkkopalvelun käyttäjälle selkeästi jos käytöstä kerätään tietoa. Tämä voidaan tehdä esimerkiksi verkkopalvelun käyttöehdoissa. Jos verkkopalveluun on mahdollista luoda tunnus, on seurannasta hyvä mainita rekisteröitymisprosessin yhteydessä. Käyttäjälle kerrottavia yksityiskohtia ovat hyvien periaatteiden mukaisesti:

- mitä kerätään
- miten tietoa käytetään ja
- kuinka kauan tietoa säilytetään. (Vrt. Croll & Sean, 2009, ss. 5-6)

Käytön seurannan laajuutta ja tarkkuutta voidaan joissain maissa ohjata laeilla. Lakitekniset tekijät on otettava huomioon käytön seurannan toteutuksessa ja käyttönotossa, jotta ei ajauduta ongelmiin projektien myöhemmissä vaiheissa. Hyvä huomio on, että useissa maissa yksityisyyttä koskeva lainsäädäntö on Internetin myötä murrosvaiheessa. Esimerkiksi Saksassa käytön seurantatietoon ei ole ennen saanut tallentaa käyttäjän IP-osoitetta, koska tämän on nähty olevan käyttäjän henkilökohtainen tieto. Tässä tapauksessa käytön seurantatietoon liittyvä laki muuttui avoimemmaksi, mutta tilanne voi yhtä hyvin olla päinvastainen, jolloin seurantajärjestelmä ja sen keräämä tieto voi muuttua laittomaksi. (Croll & Sean, 2009, ss. 5-6; *German court says IP addresses in server logs are not personal data*, 2008)

Verkkopalvelujen, jotka tallentavat käyttäjään liittyviä tietoja, tulee Suomen henkilötietolain (Henkilötietolaki (523/99) 10 §) mukaan laatia henkilörekisteristä rekisteriseloste. Rekisteriseloste on pidettävä verkkopalvelun käyttäjien saatavilla. Rekisteriselosteesta tulee ilmetä

1. rekisterinpitäjän ja tarvittaessa tämän edustajan nimi ja yhteystiedot,

2. henkilötietojen käsittelyn tarkoitus,
3. kuvaus rekisteröityjen ryhmästä tai ryhmistä ja näihin liittyvistä tiedoista tai tietoryhmistä,
4. mihin tietoja säännönmukaisesti luovutetaan ja siirretäänkö tietoja Euroopan unionin tai Euroopan talousalueen ulkopuolelle sekä
5. kuvaus rekisterin suojauksen periaatteista.

Verkkopalvelujen käytön seurantatiedon tallennus on henkilötietolain piirissä niiltä osin kun käyttäjistä tallennetaan heitä yksilöiviä tietoja. Sovellustason käytön seurantajärjestelmien keräämä tieto on etenkin keräysvaiheessa usein sellaista, josta yksittäiset käyttäjät voidaan yksilöidä. Tällöin verkkopalvelun rekisteriselosteessa tulee kuvata mitä tietoja käyttäjistä tallennetaan, miksi näitä tietoja kerätään, kenellä niihin on oikeus ja miten niitä käytetään hyväksi. (Vrt. *Henkilötietolaki*, 1999)

3 Seurantajärjestelmän toteutus verkkopalvelussa

Käytön seurantajärjestelmien toteuttamiseen verkkopalveluissa on useita tapoja joita käsiteltiin luvussa 2. Verkkopalvelun käyttöä voidaan esimerkiksi seurata ulkoisilla välineillä perustuen sivulatauksiin liittyviin HTTP-pyyntöihin. Verkkopalvelun kehittäjällä on mahdollisuus toteuttaa verkkopalvelun käytön seuraamiseen myös omia sovelluksia. Näiden verkkopalvelun sisään tai suoraan yhteyteen toteuttavien seurantajärjestelmien etuna on, että ne pääsevät kattavammin käsiksi verkkopalvelun osiin.

Toteutettaessa verkkopalvelun käytön seurantajärjestelmää tulee tuntea verkkopalvelun käyttämä arkkitehtuuri. Arkkitehtuuri vaikuttaa siihen millä tavoin tieto kulkee verkkopalvelun osien välillä. Käytön seurantajärjestelmän on kyettävä etsimään käyttötietoa verkkopalvelun oikeista osista. Verkkopalvelun käyttämä arkkitehtuuri määrittelee millä tavalla verkkopalvelun käytön seurantajärjestelmä tulee toteuttaa.

Käytön seurannan osalta järjestelmän toiminta voidaan jakaa osakokonaisuuksiin. Ensinnäkin voidaan seurata verkkopalveluun kohdistuvien sivulatausten virtaa. Sivulataustietoihin voidaan liittää verkkopalveluun liittyviä tietoja. Tällaisia tietoja ovat muun muassa onko sivun ladannut käyttäjä kirjautunut palveluun tai liittyykö sivulataus johonkin tietosisältöön verkkopalvelussa. Esimerkki verkkopalvelussa olevasta tietosisällöstä, joka voidaan yhdistää sivulataustietoon on verkkopalvelussa oleva käyttäjäprofiili. Tällöin voidaan muun muassa vertailla mikä verkkopalvelussa olevista käyttäjäprofileista on sivulatausten määrän mukaan suosituin. Tietosisältöjä on mahdollista seurata niiden lukumäärän mukaan. Esimerkiksi verkkopalveluun rekisteröityneiden käyttäjäprofilien määrän kehitystä on mahdollista tutkia. Verkkopalvelussa olevien tietosisältöjen seuranta on usein mahdotonta verkkopalvelua ulkoapäin seuraavilla järjestelmillä, koska tietosisällöt tallennetaan tietokantaan, johon pääsy on rajoitettu.

Sivulatausten seurantaan liittyy kiinteästi verkkopalvelussa suoritettujen tapahtumien seuranta. Kun verkkopalvelun käyttäjä suorittaa verkkopalvelussa sivulatauksen, saattaa tähän sivulataukseen liittyä jokin tapahtuma. Tapahtumalla tarkoitetaan verkkopalvelussa olevan toiminnon suorittamista. Yleisemmin tapahtumalla voidaan käsittää tarkoittavan kahden sivulatauksen välistä yhteyttä. Tapahtuma voi tässä tapauksessa olla esimerkiksi siirtymä kahden näkymän välillä. Oleellista tapahtumalle on, että se määrittelee verkkopalvelussa suoritettavan käytön seurannan kannalta kiinnostavan kokonaisuuden.

Tapahtumien seuranta voidaan käsittää laajemmin niin, että seurataan peräkkäisten tapahtumien sarjoja, joka mahdollistaa tapahtumien laajemman tarkastelun. Tapahtumasarjojen perusteella voidaan verkkopalvelussa pyrkiä osoittamaan tapah-

tumien syy-seuraus -suhteita. Tietyt tapahtumat verkkopalvelussa johtavat usein käyttäjien tietynlaiseen toimintaan. Syy-seuraus -suhteiden seurannan avulla on mahdollista yrittää ymmärtää miksi ja miten verkkopalvelun käyttäjät verkkopalvelua todellisuudessa käyttävät. Syy-seuraus -suhteiden seuraaminen verkkopalvelussa on erittäin haastavaa ja vaatii paljon analysointia itse verkkopalvelun toiminnasta. Usein syy-seuraus -suhteiden selvittäminen jätetään käytön seurantajärjestelmän ulkopuolelle niiden monimutkaisuuden vuoksi ja selvitetään syitä ihmisten käyttäytymiseen esimerkiksi käyttäjäkyselyiden avulla. (Vrt. Croll & Sean, 2009, ss. 53-56)

Sosiaalisten verkkopalvelujen viime aikaisen yleistymisen myötä verkkopalvelujen sosiaalisen käytön seuraamisen tarve on kasvanut. Sosiaalisissa verkkopalveluissa käyttäjät toimivat enemmän keskenään kuin perinteisissä verkkopalveluissa, joissa toiminta tapahtuu käyttäjän ja verkkopalvelua ylläpitävän organisaation välillä. Käyttäjien välinen vuorovaikutus asettaa käytön seurantajärjestelmälle tiettyjä erityisvaatimuksia, jotka tulee ottaa huomioon järjestelmää toteutettaessa. Verkkopalveluissa tapahtuva käyttäjien välinen vuorovaikuttaminen vaikuttaa siihen mitä tietoa verkkopalvelun käytöstä voidaan kerätä. (Vrt. Silius et al., 2009, s. 1; vrt. Silius et al., 2010, s. 3)

Käytön seurantajärjestelmän toteuttamiseen liittyy kiinteästi miten kerätyt tiedot tallennetaan ja miten tietoa voidaan käsitellä jälkikäteen. Käytön seurantatiedon viennin mahdollisuus verkkopalvelusta muihin järjestelmiin on oleellinen osa hyvää käytön seurantajärjestelmää. Lisäksi verkkopalvelun käytön seurantajärjestelmän toteutukseen liittyy kerätyn käytön seurantatiedon visualisointi. Seurantatiedon visualisointia kuvataan luvussa 4.

3.1 Verkkopalveluiden arkkitehtuuri

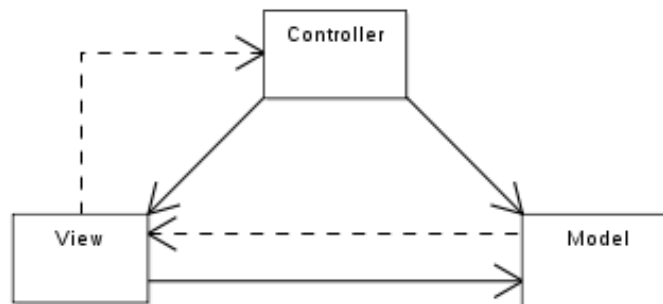
Verkkopalvelut toimivat aina jonkinlaisen arkkitehtuurin mukaisesti. Arkkitehtuuri määrittelee sen miten verkkopalvelu rakentuu ja miten verkkopalvelun osat keskustelvat keskenään. Mitä selkeämmin verkkopalvelun loogiset osat ovat eriytetty toisistaan, sitä helpompaa on verkkopalvelun käyttötietojen kerääminen. Kun verkkopalvelun arkkitehtuuri on hierarkkinen ja yhtenäinen, toimii verkkopalvelu aina samalla tavalla samankaltaisissa tilanteissa. Tämä yksinkertaistaa verkkopalvelun käytön seurantajärjestelmän toteuttamista, koska verkkopalvelussa liikkuvaa tietoa käsittelevät osat ovat yksikäsitteisiä. Verkkopalvelun ja toisaalta verkkopalvelun seurantajärjestelmän toteutus on hyvä suunnitella aina jonkin arkkitehtuurin mukaisesti, koska se yksinkertaistaa ja jäsentää toteutustyötä ja tekee siitä ulkopuolisen näkökulmasta helpommin ymmärrettävämmän.

MVC-arkkitehtuuri (Model-View-Controller architecture) on yleinen ja usein

ohjelmistotuotannossa hyödynnetty ohjelmistoarkkitehtuuri, jossa järjestelmän käyttöliittymä erotetaan sovelluksen toiminnallisuudesta. MVC-arkkitehtuuri on teoreettinen malli, joka kuvaa korkealla tasolla verkkopalvelun osat ja niiden väliset suhteet. MVC-arkkitehtuurin osat ovat:

- malli (model)
- näkymä (view) ja
- ohjain (controller).

Arkkitehtuurin mukaisesti malli huolehtii tiedon tallentamisesta, ylläpidosta ja käsittelystä. Näkymän tehtävä on esittää mallin ja ohjaimen tarjoamat resurssit käyttäjälle. Ohjain taas käsittelee näkymältä tulevia käskyjä ja käsittelee mallia niiden mukaisesti. Kuvassa 3.1 nähdään arkkitehtuurin osien väliset keskusteluyhteydet ja yhteyksien suunnat. Kuvassa yhtenäinen viiva kuvaa mallin osien välistä suoraa yhteyttä (association) ja katkoviiva osien välistä epäsuoraa yhteyttä. Yhteydellä tarkoitetaan, että arkkitehtuurin osa pystyy vaikuttamaan toisen osan toimintaan. Epäsuorassa yhteydessä arkkitehtuurin osien välinen vaikutussuhde on välillinen.



Kuva 3.1: MVC-arkkitehtuurin rakenne ja osien väliset yhteydet.

Verkkopalveluissa MVC-arkkitehtuuria tai sen variaatiota hyödynnetään yleisesti. Arkkitehtuurin voidaan nähdä toimivan verkkopalveluissa aina vähintään niin, että ohjain on verkkopalvelussa käytetty ohjelmointikieli kuten Python tai PHP (Hypertext Preprocessor) ja mallina verkkopalvelun käyttämä tietokanta kuten MySQL tietokannan hallintajärjestelmä. Ohjain tuottaa mallin avulla käyttäjille näkymiä ja näkymät muodostuvat esimerkiksi HTML-merkkauksen avulla. Verkkopalvelujen kehitysympäristöt, kuten CodeIgniter³², Django³³ ja Drupal³⁴, toteuttavat MVC-mallin hierarkkisen rakenteen niin, että arkkitehtuurin määrittelemät verkkopalvelun osat on lisäksi tiedostotasolla eroteltu toisistaan. (Vrt. Burbeck, 1987)

³²<http://codeigniter.com/>

³³<http://www.djangoproject.com/>

³⁴<http://drupal.org/>

3.2 Sivulatausten seuranta

Verkkopalvelun sivulatausten seuraaminen voidaan nähdä verkkopalvelun käytön seurannan perustana. Sivulataus on yksinkertaisin tietoalkio, joka verkkopalvelun käytöstä voidaan tunnistaa. Sivulatauksia voidaan seurata koko sivuston laajuisesti, kohdistuen yksittäiseen tietosisältöön tai millä vain halutulla tavalla. Jos seurattava verkkopalvelu on toteutus keskustelualueesta (forum), voidaan sivulatauksia seuraamalla kertoa esimerkiksi montako kertaa keskustelu-alueen aihetta “x” on luettu. Huomioitavaa on, että tällainen tieto voidaan välittää verkkopalvelun käyttäjälle tietona: tämä sivu on luettu “n” kertaa (kuva 3.2). Käytön seurantatietoa voidaan jakaa käyttäjien suuntaan käyttäjien käyttökokemuksen (user experience, UX) parantamiseksi.

Ketju / Ketjun aloittaja	Arvio	Viimeisin ↓	Vastausta	Luettu
↓ Nosto: "Mikä kannettava minulle?" -yleisthreadi (1 2 3 ... Viimeinen sivu)		Eilen 13:48 ↓	831	55448

Kuva 3.2: Käytön seurantatietoa voidaan käyttää rikastamaan verkkopalvelun käyttäjän käyttökokemusta, kuten nähdään tässä keskustelualueen säikeen kohdalla.

Sivulatausten seurannan toimintaperiaate on tallentaa jokainen verkkopalveluun tuleva sivulatauspyyntö, kerätä pyyntöön liittyvät tiedot ja tallentaa ne. Käytön seurantajärjestelmän tulee sijoittua hierarkkisesti verkkopalvelussa, että se havaitsee sivulatauspyynnöt riippumatta sivulatauspyynnön mukana välitettävistä tiedoista. Jos esimerkiksi käyttäjän pyytämä sivu vaatii, että käyttäjä on kirjautunut verkkopalveluun ja järjestelmä uudelleenohjaa pyynnön kirjautumislomakkeelle, tulee seurantajärjestelmän pystyä rekisteröimään nämä molemmat sivulatauspyynnöt.

Sivulatausten seuraamisessa oleellista on kerätä kaikki pyyntöön liittyvä tieto ja tallentaa se. Yleisiä verkkopalvelusta kerättäviä tietoja ovat sivulatauksen aikaleima, asiakasohjelman tiedot ja se mikä oli käyttäjän edellinen sivulatauspyyntö. Nämä tiedot ovat kerättävissä verkkopalvelun ulkopuolisilla seurantajärjestelmillä. Verkkopalvelua sisältäpäin seuraavalla järjestelmällä on mahdollista liittää sivulataustietoihin verkkopalveluun liittyvää tietoa. Käytön seurantatietoon on esimerkiksi mahdollista liittää tieto siitä, onko sivulatauksen tehnyt käyttäjä kirjautunut verkkopalveluun. Tämä on mahdollista, koska seurantajärjestelmällä, joka kytkeytyy suoraan verkkopalveluun, on oikeus käyttää verkkopalvelun osia laajemmin kuin ulkopuolisilla järjestelmillä.

Käytön seurantajärjestelmän toteuttajalla tai toteuttajilla tulee olla kattava näkemys verkkopalvelusta johon seurantajärjestelmä toteutetaan. Verkkopalvelun sisäisen käytön seurantajärjestelmän täytyy tuntea kaikki verkkopalvelun osat,

joita sen halutaan seuraavan. Etenkin sellaiset verkkopalvelun osat, jotka eivät ole verkkopalvelujen kesken yleisesti yhteisiä, täytyy erikseen esitellä käytön seurantajärjestelmälle. Jos verkkopalvelussa esimerkiksi käytetään erilaisten tietosisältöjen nimeämiseen avainsanoja, on käytön seurantajärjestelmän pystyttävä tunnistamaan ne verkkopalvelun osat, joissa avainsanoja (tags) käytetään, jotta se voi tallentaa niihin liittyvää tietoa. Kattavan sivulatauksia seuraavan käytön seurantajärjestelmän toteuttaminen vaatii, että seurantajärjestelmä tuntee kaikki seurattavana olevat tekijät.

Verkkopalvelujen osien tunnistaminen vaatii, että niiden toiminnot ja ominaisuudet on määritelty. Määrittelyn perusteella seurantajärjestelmä on mahdollista toteuttaa kattavaksi. Sivulatausten seurantatietojen kattavuudessa voi syntyä ongelmia jos verkkopalvelu on hyvin laaja tai jos seurantajärjestelmää toteuttava taho ei tunne verkkopalvelun toimintaa tarpeeksi kattavasti. Verkkopalvelun ominaisuuksien ja toiminnallisuuden dokumentointi onkin hyvin tärkeää, kun verkkopalveluun halutaan toteuttaa uusia ominaisuuksia kuten seurantajärjestelmiä verkkopalvelun elinkaaren myöhemmissä vaiheissa. Kattavien verkkopalvelun määrittely- ja suunnitteludokumenttien perusteella verkkopalvelun toiminnallisuuteen voidaan tutustua. Dokumentoinnin perusteella on mahdollista varmistaa, että kaikki verkkopalvelun osat tulevat seuratuiksi.

3.2.1 Sivulatausten suodattaminen

Sivulatausten suodatus tarkoittaa, että verkkopalveluun kohdistuvista sivulatauspyynnöistä rajataan määritellyillä ehdoilla halutut sivulatauspyynnöt käsiteltävän joukon ulkopuolelle. Rajaus on mahdollista tehdä joko tietojen tallennusvaiheessa tai tietojen käsittelyvaiheessa. Sivulataustietojen suodattaminen on tärkeää, koska on mahdollista, että suuri osa verkkopalveluun kohdistuvista sivulataustapahtumista tulee asiakasohjelmilta, joista verkkopalvelun käyttöä seuraava taho ei ole kiinnostunut. Tähän joukkoon voivat kuulua vihamieliset hyökkäysryitykset (denial of service, DOS), roskapostitusrobotit (spambot) ja hakukonerobotit (web crawler). Tavanomaisesti käytön seurantatiedosta halutaan suodattaa verkkopalvelun kehittäjien tuottama liikenne, koska heidän toimiansa näkyminen esimerkiksi visualisoinneissa saattaisi vääristää tuloksia.

Sivulatausten suodatus voidaan tehdä kahdessa vaiheessa. Jos halutaan vähentää tiedon jälkikäsitteilyä ja pienentää lokitiedon viemää levytilaa, on mielekästä suodattaa ei-halutut verkkopalvelua käyttävät kohteet jo tiedon tallennusvaiheessa. Tietojen tallentamatta jättäminen tarkoittaa kuitenkin, ettei tällöin käytön seurantatietoa näistä asiakasohjelmista ole saatavilla. Saattaa olla mahdollista, että verkkopalvelun elinkaaren aikana syntyy jossain vaiheessa tarve tietää miten tajaan verkkopalvelua indeksoidaan hakukonerobottien toimesta.

Suodatus voidaan suorittaa tietojen visualisoinnin yhteydessä. Tällöin ei-halutut kohteet määritellään visualisointia luotaessa, joka mahdollistaa ei-haluttujen kohteiden mukautuvamman uudelleenmäärittelyn. Ei-haluttujen kohteiden suodatus visualisoinnin yhteydessä kuitenkin hidastaa visualisoinnin luontia, koska prosessoitavaa tietoa on enemmän. On mahdollista, että jopa 60 %:a verkkopalveluun kohdistuvista sivulatauspyynnöistä tapahtuu erilaisten robottiohjelmien toimesta. Huomioitavaa on, että mitä suositumpi verkkopalvelu on kyseessä, sitä suurempi osuus verkkopalveluun kohdistuvasta liikenteestä tulee muista lähteistä kuin todellisilta verkkopalvelun käyttäjiltä. (Vrt. Croll & Sean, 2009, ss. 129-130)

3.2.2 Verkon ulkopuolella tapahtuvan käytön seuraaminen

Internetissä on paljon sisältöjä, jotka ladataan verkkopalvelusta, mutta jotka luetaan verkon ulkopuolella (offline). Esimerkiksi PDF-dokumentit (Portable Document Format) ja PowerPoint-esitykset ovat verkkopalvelussa olevia sisältöjä, jotka ladataan käyttäjän paikalliselle (local) tietokoneelle ja luetaan erillisellä ohjelmalla. Näiden sisältöjen käytöstä ei perinteisellä verkkopalvelun käytön seurantajärjestelmällä saada muuta tietoa kuin, että tiedosto on ladattu palvelimelta. Siitä miten tietoa on käytetty ja mitä sivuja esimerkiksi PDF-dokumenteissa on selattu jää seurannan ulkopuolelle. Näiden sisältöjen käytön seurantatieto voi kuitenkin olla tietyissä tapauksissa ja tietyissä verkkopalveluissa ensiarvoisen tärkeää.

Verkon ulkopuolella luettavien sisältöjen tapauksessa on tarpeen, että lukuohjelmaan toteutetaan sisällön käytön seurantajärjestelmä, joka kerää tiedot väliaikaiseen säiliöön (välimuisti, cache) ja havaitessaan Internetyhteyden, lähettää kerätyt tiedot välimuistista seurantatietoa keräävään palveluun. Kun seurantatieto on onnistuneesti kerätty, voidaan tiedot poistaa lukuohjelmasta.

Ongelmana seurannan toteuttamisessa lukuohjelmaan on usein, että ohjelmat joilla tiedostoja luetaan, eivät ole verkkopalvelun hallinnoijan toteuttamia. Tästä johtuen seurantajärjestelmän lisääminen lukuohjelmaan voi olla vaikeaa tai mahdotonta. Usein oman lukuohjelman ja tiedostomuodon luominen tätä tarkoitusta varten ei ole järkevää. Paras tilanne on, että tiedostomuoto on luettavissa vain yhdellä tietyllä ohjelmalla, johon on mahdollista toteuttaa ulkopuolisia lisäosia, joiden avulla seuranta tehtäisiin.

Verkon ulkopuolisen sisällön seurannan toteutus rajataan kuitenkin vahvasti tämän työn ulkopuolelle ja pääpaino on verkossa tapahtuvan käytön seuraamisessa. Verkon ulkopuolisten sisältöjen seuraamisen tarpeen tiedostaminen on kuitenkin oleellista verkkopalvelujen käytön seurannan kokonaiskuvan hahmottamisen kannalta. (Vrt. Kaushik, 2010, ss. 248-249)

3.2.3 Vaihtoehtoiset sisällön lukutavat

Verkkopalvelujen käytön seuranta pohdittaessa on hyvä huomioida, että verkkopalveluissa olevaa tietosisältöä on mahdollista lukea eri teknologioiden avulla. Normaalisti verkkopalveluissa olevaa sisältöä luetaan suoraan verkkopalvelusta. Sisällön lukemiseen verkkopalvelussa saattaa kuitenkin olla muita keinoja. Esimerkiksi RSS-virta (Really Simple Syndication tai RDF (Resource Description Framework) Site Symmary) mahdollistaa verkkopalvelussa olevan sisällön lukemisen ilman verkkopalvelun selaamista. Vaihtoehtoiset sisällön lukutavat vaikuttavat verkkopalvelun käyttöön niin, että ne pienentävät verkkopalveluun kohdistuvan liikenteen määrää. Jos vaihtoehtoisia sisällön lukutapoja ei oteta verkkopalvelun seurannassa huomioon, saatetaan tehdä virheellisiä johtopäätöksiä verkkopalvelun käytön suosiosta. Vaihtoehtoiset verkkopalvelun lukutavat eivät siis tuota verkkopalvelun käytön seuraimen kannalta samalla tavalla sivulataustapahtumia kuin normaali verkkopalvelun käyttö. Vaihtoehtoisten lukumuotojen seuranta on kuitenkin mahdollista, koska niiden lataaminen palvelimelta, jolla tietosisältö sijaitsee, voidaan havaita ja tallentaa. (Vrt. Kaushik, 2010, ss. 150-151)

3.3 Seurattavien tietojen määrittely

Käytön seurantajärjestelmää toteutettaessa tulee määritellä mitä tietoja verkkopalvelusta halutaan seurata. Seurattavat asiat tulee nimetä ja niille tulee määritellä miten niitä seurataan. Seurattavia tietoja määriteltäessä on hyvä pohtia miten tietoja tullaan myöhemmin hyödyntämään ja visualisoimaan. Käytön seuraamista pohdittaessa on tarpeellista nähdä toteutus läpi koko projektin. Käytön seurantatiedon visualisointi on huomattavasti yksinkertaisempaa, kun visualisointiin liittyvät tekijät on otettu huomioon kerättävien tietojen määrittelyssä. Pohdittaessa kerättävien tietojen merkitystä ja tiedon avulla mahdollisesti saavutettavia tavoitteita, voidaan havaita verkkopalvelusta täysin uusia seurattavia asioita, joita ei välttämättä muuten huomattaisi ottaa seurantaan mukaan. Seurattavien tietojen määrittelyä voidaan lähestyä järjestelmän ominaisuuksien kautta, jolloin kirjataan ylös mitä toimintoja verkkopalvelussa on ja varmistetaan, että kaikista niistä kerätään tietoa. Ongelmaa on mahdollista lähestyä ratkaisulähtöisesti, jolloin mietitään mitä käytön seurantatiedolla halutaan tehdä.

Verkkopalvelusta kerättävät tiedot voidaan määritellä olemassa olevan tiedon näkökulmasta tai tiedon käytön näkökulmasta. Lähestyttäessä kerättävien tietojen määrittelyä olemassa olevan tiedon näkökulmasta, etsitään järjestelmästä kaikki olemassa oleva tieto ja toteutetaan niiden keräys. Vaihtoehtoinen lähestymistapa on pohtia, mitä käytön seurantatietoa järjestelmästä halutaan kerätä. Tällöin määritellään halutut tiedot ja toteutetaan niiden kerääminen verkkopalveluun. Lähestymis-

tavoissa on molemmissa hyvät ja huonot puolensa ja haluttaessa kerätä tietoa mahdollisimman hyvin, tulisi pyrkiä tarkastelemaan tietojen määrittelyä kummastakin näkökulmasta. Jos verkkopalvelusta kerätään kaikki olemassa oleva tieto pohtimatta miten tätä tietoa aiotaan käyttää, saatetaan joutua tilanteeseen, jossa kerättyä tietoa ei pystytä hyödyntämään. Tällöin on mahdollista, että suurin osa kerätystä tiedosta ei tule hyödynnetyksi millään tavalla, jolloin osa keräysjärjestelmästä on toteutettu turhaan. Pohdittaessa vain miten tietoa tullaan käyttämään voi toisaalta käydä niin, että jokin oleellinen verkkopalvelun osa jää seurannan ulkopuolelle.

Lähestymistapojen huomioiminen on riippuvainen verkkopalvelusta sekä ihmisten taustoista, jotka seurantajärjestelmää ovat toteuttamassa. Ihmiselle on luontevaa lähestyä ongelmaa siitä lähtökohdasta, joka on hänelle itselleen luontevin. Luonteva lähestymistapa on usein verrannollinen ihmisen koulutukseen ja osamiseen. Tämä lähestymistapa saattaa kuitenkin johtaa ongelmiin jos seurantajärjestelmää toteuttavilla henkilöillä ei ole käsitystä joltain käytön seurantajärjestelmän toteutuksen osa-alueelta. Käytön seurantatietoa kerätessä tulisi esimerkiksi välttää sellaisia tilanteita, joissa verkkopalvelun käytöstä tallennettua tietoa ei pystytä hyödyntämään käytettävissä olevilla tiedon visualisointijärjestelmillä. Tämänkaltainen tilanne voi esiintyä jos seurantajärjestelmä on toteutettu sellaisesta teknisestä näkökulmasta, joka estää seurantatiedon viennin visualisointijärjestelmän käsiteltäväksi. Seurattavien tietojen määrittelyssä tulisi pyrkiä näkemään verkkopalvelun käytön seuranta kokonaisvaltaisena prosessina, jossa kaikki seurannan vaiheet pyritään ottamaan huomioon toteutuksen edetessä. (Vrt. Haikala & Märijärvi, 1995/2002, s. 25)

Seurattavia tietoja määriteltäessä voidaan nostaa esille kysymys “kerätäänkö verkkopalvelun käytöstä kaikki tieto” ja toisaalta “onko kerätty tieto todella sitä mitä sen luullaan olevan”. Etenkin monipuolisten ja sen myötä usein monimutkaisten verkkopalvelujen tapauksessa on käytön seurantajärjestelmän toteutus niin laaja, että sen hallitseminen on vaikeaa. Tämän johdosta on tärkeää, että kerättyä seurantatietoa pystytään koekäyttämään jo prosessin alkuvaiheissa ja siten todentamaan kerätyn tiedon kattavuus sekä onko kerätty tieto todella sitä mitä sen halutaan olevan.

3.3.1 Neljä suurta kysymystä

Croll & Sean (2009, ss. 53-56) mukaan etsimällä vastaukset neljään suureen kysymykseen pystytään määrittelemään mihin käytön seurantatiedon keruussa tulee keskittyä. Kysymykset ovat käyttäjälähtöisiä ja niiden määrittellään olevan:

1. mitä käyttäjät tekevät verkkopalvelussa
2. miten käyttäjät toimivat verkkopalvelussa

3. miksi käyttäjät toimivat verkkopalvelussa ja

4. onko käyttäjien mahdollista tehdä verkkopalvelussa se mitä haluavat?

Kysymysten asettelu on hyvä, koska niissä tarkastellaan verkkopalvelun käytön seurantaa käyttäjälähtöisesti. Vastausten pohtiminen pakottaa verkkopalvelun kehitysryhmän asettumaan käyttäjän asemaan ja näkemään käytön seuraamisen käyttäjän näkökulmasta. Ensimmäisen kysymyksen avulla saadaan tietoa siitä mitä järjestelmällä voidaan tehdä. Kysymyksen perusteella saadaan periaatteessa lista verkkopalvelun toiminnoista. Listalla olevat toiminnot on kuitenkin pohdittu käyttäjälähtöisesti, joten osa toiminnoista, jotka voidaan määritellä teknisesti, saattavat esimerkiksi yhdistyä käyttäjän näkökulmasta tarkasteltuna yhdeksi suuremmaksi toiminnallisuudeksi. Kysymystä pohdittaessa on hyvä huomioida, että seurantatietoa syntyy verkkopalveluissa monilla tavoilla. Useat verkkopalvelut hyödyntävät toiminnoissaan nykyaikaisia teknologioita kuten AJAX (Asynchronous JavaScript And XML) tai DHTML (Dynamic HTML), joita käyttäessään käyttäjän ei tarvitse suorittaa uutta sivulatauspyyntöä. Tämä tarkoittaa, että käyttäjän on mahdollista toimia verkkopalvelussa ilman, että hän siirtyy sivulta toiselle.

Toisella kysymyksellä pyritään etsimään vastausta siihen millä tavalla verkkopalvelun käyttäjät käyttävät verkkopalvelua. Kysymys on käytettävyytlähtöinen, joka on itsessään oleellinen osa hyvää verkkopalvelua. Nykyaikaisissa verkkopalveluissa esimerkiksi toimintojen käyttöpolut saattavat poiketa käyttäjien kesken. Vastamalla kysymykseen voidaan havaita, että pyrkiessä johonkin verkkopalvelun osaan päätyy usea käyttäjä johonkin toiseen, väärään verkkopalvelun osaan matkalla haluamaansa sijaintiin. Havaitsemalla tämänkaltainen tilanne voidaan prosessia selkiyttää ja näin parantaa verkkopalvelun käytettävyyttä. Verkkopalvelun seurantatiedon keräyksen kannalta Croll & Sean määrittelemä toinen kysymys auttaa pohtimaan, mitä tietoa verkkopalvelun käytöstä halutaan kerätä. (Croll & Sean, 2009, ss. 53-56)

Tietoa siitä miksi käyttäjät toimivat verkkopalvelussa kuten toimivat on vaikea selittää käytön seurantajärjestelmän avulla. Käytön seurantatiedon avulla on mahdollista näyttää, että käyttäjät toimivat jollakin tavalla, mutta käyttäytymisen selittäminen on usein käytön seurantajärjestelmän avulla vaikeaa. Erilaisten käyttötapausten avulla voidaan osoittaa, että käyttäjät toimivat jonkin prosessimallin mukaisesti, mutta on vaikeaa määritellä miksi käyttäjät toimivat prosessin kuvaimalla tavalla. Kun voidaan osoittaa, että käyttäjät kirjautuvat verkkopalveluun usein vierailtuaan jonkun toisen käyttäjän profilissa, on mahdollista yrittää sanoa, että käyttäjän profiilinäkymä on sellainen, joka motivoi käyttäjiä kirjautumaan. Näin voidaan väittää, koska useat käyttäjät kirjautuvat verkkopalveluun vierailtuaan ensin jonkun toisen käyttäjän profilissa. Todellisten syiden osoittaminen

ihmisten käyttäytymisessä vaatii kuitenkin usein laajempaa tiedon analysointia. Internetissä kysymykseen “miksi” pyritään etsimään vastausta käyttäjäkyselyiden avulla. Käyttäjäkyselyt toteutetaan usein verkkokyselyinä ja ne ovat hyvin tehtyinä tehokas tapa kerätä tietoa, koska kerätty tieto tulee suoraan käyttäjiltä. Käyttäjäkyselyitä voidaan tehdä verkkopalveluissa koskien koko verkkopalvelua tai vain osaa siitä. (Kaushik, 2010, ss. 179-187; Croll & Sean, 2009, ss. 53-56)

Neljäs kysymys käsittelee verkkopalvelun käytön seuranta saavutettavuuden näkökulmasta. Kysymyksen vastauksena saadaan tietoa verkkopalvelun suorituskyvystä ja käytetyistä teknologioista. Esimerkiksi voidaan saada selville, että jokin verkkopalvelun osa on toteutettu käyttäen sellaista teknologiaa, jota useiden verkkopalvelujen käyttäjien päätelaitteet tai asiakasohjelmat eivät tue. (Croll & Sean, 2009, ss. 53-56)

Huomioitavaa on, että Croll & Sean (2009) kuvaama malli on tehty etenkin yrityksen näkökulmasta katsoen ja käyttäjä nähdään pääasiassa asiakkaana (customer). Tämä tarkoittaa, että kun mallia hyödynnetään tulee sen tulokinnassa muistaa, että mallin tarkoitus on lähtökohtaisesti tuottaa hyödyntäjälleen vain mahdollisimman suuri liikevoitto. Jos käytön seuraamisen tavoitteena on esimerkiksi tutkia verkkopalvelun toimintaa ja ihmisten käyttäytymistä tieteellisistä lähtökohdista on hyvä pyrkiä tarkastelemaan käytön seurantatiedon määrittelyä useampien näkökulmien kautta.

3.3.2 Kaistan käytön seuranta

Verkkopalvelun käytön seurannan mittarina voidaan käyttää verkkopalvelusta ladatun tiedon määrää, jota voidaan seurata palvelimelta lähetettyjen tavujen (bytes) muodossa. Verkkopalvelujen yksittäisten näkymien koko on tyypillisesti 1 – 100 kilotavua, mutta esimerkiksi multimediatiedostot (media files) voivat olla kooltaan useiden megatavujen kokoisia. (Vrt. Blaauw et al., 1962, s. 75-79)

Kun käyttäjän asiakasohjelma lataa verkkopalvelun palvelimelta tiedostoja, näkyy se palvelimella tiedon lähettämisenä ja tieto lähetyn tiedon määrästä on mahdollista tallentaa. Verkkopalvelun kaistan käytön seuranta voi antaa erilaista kuvaa verkkopalvelun käytöstä kuin sivulatausmääriin perustuva käytön seurantatieto. Jos verkkopalvelussa on kooltaan hyvin suuria tietosisältöjä, mutta niiden lukumäärä on pieni, näyttää sivulatauksiin perustuva käytön seurantatieto siltä kuin verkkopalvelua hyödynnettäisiin vain vähän. Seurattaessa verkkopalvelun käyttämää kaistaa havaitaan tämänkaltaisen verkkopalvelun käyttö. Virta-sisällöt (stream) ovat esimerkiksi sen tyyppisiä, että niiden määrä verkkopalvelussa on usein vähäinen, mutta niiden aiheuttaman liikenteen määrä on suuri. Verkkopalvelun kaistan käytön seuranta voidaan käyttää mittarina, jonka avulla voidaan arvioida tulisiko verkkopalvelun käytössä olevaa kaistaa mahdollisesti laajentaa.

Useissa palvelinympäristöissä verkkopalvelun kaistan käytön seuranta voidaan toteuttaa esimerkiksi MRTG-työkalulla³⁵ (Multi Router Traffic Grapher), joka on ilmainen kaistan käytön visualisointityökalu. Verkkopalvelun kaistan käytön seuranta on mahdollista sisällyttää osaksi verkkopalvelun sisäistä käytön seurantajärjestelmää, mutta MRTG-työkalu analysoi ja tallentaa käytön seurantatietoa ilman verkkopalvelukohtaista toteutustyötä, joten sen hyödyntäminen tähän tarkoitukseen voi olla mielekästä.

3.3.3 Palvelinkuorman seuranta

Verkkopalvelu on sijoitettu yhdelle tai useammalle palvelimelle. Palvelinkoneella pääosa työstä tehdään hyödyntäen verkkopalvelun suoritinta ja keskusmuistia. Näiden fyysisten osien käytön seurannalla voidaan havaita onko käytössä olevalla palvelimella tai palvelimilla liian vähän kapasiteettia verkkopalvelun käyttäjämäärien palvelemiseen. Keskusmuistin tai suorittimen kapasiteetin vähyys näkyy käyttäjälle verkkopalvelun toimintojen suoritusajan kasvuna. Etenkin raskaampien verkkopalvelujen ominaisuuksien suorittamisaika voi kasvaa niin suureksi, että niiden suoritus keskeytyy aikakatkaisun (timeout) myötä.

Suorittimen ja keskusmuistin lisäksi verkkopalvelu hyödyntää palvelimen kovalevytilaa. Vapaan ja käytetyn kovalevytilan seuranta on siinä määrin tärkeää, että jos levytila loppuu, on verkkopalveluun tämän jälkeen mahdotonta lisätä uutta tietosisältöä. Verkkopalvelun käyttämän kovalevyn ja palvelimen vapaan levytilan määrä on tästä syystä tärkeää tarkkailla. Toisaalta käytetyn kovalevytilan seurannalla voidaan mitata verkkopalvelun kokoa ja vertailla verkkopalvelua muihin verkkopalveluihin. Verkkopalvelun käyttämän kovalevyn tilaa on mahdollista seurata käyttäjäkohtaisesti. Jos verkkopalvelujen käyttäjien on mahdollista lisätä verkkopalveluun tietosisältöä, on tämän tietosisällön määrää mielekästä seurata.

Palvelin kapasiteetin seuranta voidaan toteuttaa useimmissa palvelin ympäristöissä kuten kaistan käytön seurantaakin MRTG-työkalulla. MRTG-työkalu tallentaa ja visualisoi verkkopalvelun palvelimen käyttämän kapasiteetin määrää ajan funktiona, joten sen avulla on voidaan seurata verkkopalvelun palvelin tarpeen kehitystä.

3.3.4 Asiakaslaitteiden monikäyttö

Kerätessä tietoa verkkopalvelun käytöstä voidaan usein vain todeta, että tietty asiakasohjelma tai joku verkkopalvelun käyttäjä käytti verkkopalvelua. Todellisuudessa tilanne voi olla se, että verkkopalvelua käyttää samanaikaisesti useampi henkilö saman päätelaitteen välityksellä. Käytön seurantajärjestelmälle nämä käyttäjät

³⁵<http://oss.oetiker.ch/mrtg/>

näyttäytyvät yhtenä ja samana instanssina, koska tavallisesti seurantajärjestelmä ei pysty seuraamaan käyttäjän käyttöympäristöä.

Asiakaslaitteiden monikäytön merkitys verkkopalvelun seurannalle on verkkopalvelukohtaista. Toiset verkkopalvelut ovat luonteeltaan sellaisia, että niiden käyttö on luontevampaa joukossa, kun taas toisia verkkopalveluja käytetään melkein aina yksin. Esimerkki verkkopalvelusta, jonka käyttö on luontevaa joukossa, on opintojaksolla käytettävä harjoitustyömateriaali. Harjoitustyömateriaalia saattaa lukea opintojaksolla 30 kirjautunutta käyttäjää, mutta harjoitustyön voi silti suorittaa tähän suhteessa moninkertainen määrä. Tämä selittyy sillä, että harjoitustyötä tehdään ryhmissä ja materiaalia luetaan käyttäen vain yhden ryhmän jäsenen tunnusta. Toisaalta pankkien verkkopalvelut ovat esimerkki verkkopalvelusta, jota käytetään usein yksin.

Monikäyttöön liittyvät asiat tulee ottaa huomioon etenkin siinä tapauksessa kun käytön seurantatietoa tulkitaan ja analysoidaan. Käytön seurantajärjestelmän kannalta asiakaslaitteiden monikäyttöä on kuitenkin vaikea todentaa. Monikäytön tunnistaminen on mahdollista hyödyntämällä käyttäjän päätelaitteessa olevia toiminnallisuuksia. Esimerkiksi käyttäjän päätelaitteen kameratoiminnallisuudella voidaan tunnistaa päätelaitteen äärellä olevien henkilöiden määrä. Käyttäjän päätelaitteissa olevien toiminnallisuuksien käyttäminen vaatii, että käyttäjä antaa verkkopalvelulle oikeuden niiden käyttöön, mikä saattaa häiritä käyttäjän toimintaa verkkopalvelussa. Jos verkkopalvelussa on jonkin tietosisällön kohdalla erityisen tärkeää tietää kuinka moni henkilö tietosisältöä absoluuttisesti hyödyntää, voidaan tämä kysyä käyttäjiltä verkkopalvelun käytön yhteyteen sijoitettavalla lomakkeella.

3.3.5 Mobiililaitteiden seuranta

Nykyään verkkopalveluja käytetään yhä enenevässä määrin erilaisilla mobiililaitteilla. Mobiililaitteella tarkoitetaan tässä matkapuhelimia ja PDA-laitteita (Personal Digital Assistant). Mobiililaitteille on tyypillistä, että ne sisältävät erilaisia toiminnallisuuksia kuin pöytäkoneet ja niiden näyttö on kooltaan pienempi kuin pöytä tietokoneissa (mukaan lukien kannettavat tietokoneet). Mobiililaitteiden käytön seuranta asettaa käytön seurantajärjestelmälle erinäisiä haasteita. Jos käytön seurantajärjestelmä toimii käyttäen JavaScript-teknologiaa, rajaa tämä pois useat, etenkin vanhemman sukupolven mobiililaitteet, koska ne eivät tue tätä teknologiaa. Mobiililaitteiden seurantaan liittyy erityisiä seurantaan liittyviä tietoja, jotka ovat ominaisia vain mobiililaitteille. Mobiililaitteiden ja mobiilisivustojen käytön seurantaan onkin kehitetty täysin omia järjestelmiä. Esimerkiksi suomalainen QAim³⁶ on toteuttanut kaupallisen mobiilisivustojen seuranta- ja analyysijärjestelmän, jonka avulla mobiililaitteiden käyttöä verkkopalveluissa voidaan seurata (”Mobiilisivut

³⁶<http://www.qaimgroup.com/>

saavat kävijämittauksen”, 2010).

3.3.6 Paikkatiedon seuranta

Verkkopalvelua käyttävästä asiakasohjelmasta on mahdollista päätellä päätelaitteen geologinen sijainti. Sijaintitiedon määrittely voidaan tehdä perustuen päätelaitteen IP-osoitteeseen tai asiakasohjelman jakamiin paikkakoordinaatteihin (pituus- ja leveysaste). IP-paikannuksessa huomioitavaa on, että WLAN-verkoissa (Wireless Local Area Network) toimivat päätelaitteet voidaan paikantaa tarkemmin kuin LAN-verkoissa (Local Area Network) toimivat päätelaitteet. Mobiililaitteiden paikannuksessa voidaan hyödyntää päätelaitteen mahdollista GPS-toiminnallisuutta (Global Positioning System). Mobiililaitteissa paikannus on mahdollista tehdä perustuen laitteen näkemien tukiasemien signaalien voimakkuuksiin. GPS:ää avustava, esimerkiksi päätelaitteen tukiasematietoihin perustuva, teknologia tunnetaan nimellä A-GPS (Assisted GPS).

Paikannusmenetelmien paikannustarkkuudet eroavat toisistaan merkittävästi. IP-osoitteeseen perustuva paikannus pystyy määrittämään käyttäjän sijainnin kaupungin tarkkuudella, kun WLAN-verkossa oleva käyttäjä on mahdollista paikantaa noin 10 metrin tarkkuudella. WLAN-verkkojen tarkempi paikantaminen perustuu kerättyihin WLAN-signaalien sijaintitietoihin. Sijaintitietotietokantoja ovat keränneet ainakin Google ja Skyhook Wireless. Satelliittisignaaleihin perustuvan GPS-paikannuksen avulla käyttäjän sijainti saadaan määritettyä noin metrin tarkkuudella.

Paikkatiedon seurantaan on olemassa virallinen Geolocation API -standardi (Application Programming Interface), joka määrittelee miten päätelaitteiden tulee jakaa paikkatietonsa ja toisaalta miten asiakasohjelman paikkatietoa voidaan hyödyntää. Standardin avulla on mahdollista toteuttaa paikkatiedon keräämiseen perustuvia tiedon keräysjärjestelmiä. Standardin mukainen asiakasohjelman sijainnin jakava toiminnallisuus löytyy Mozilla Firefox ja Google Chrome -selainten uusimmista versioista. Sovellus varmistaa paikkatiedon jakamisen käyttäjältä ennen sen lähettämistä verkkopalvelulle. (Popescu, 2009)

3.3.7 Käyttäjän ympäristön seuranta

Käyttäjän asiakasohjelmasta voidaan kerätä käyttäjän tekniseen käyttöympäristöön liittyviä tietoja. Tällaisia tietoja ovat käyttäjän käyttämä resoluutio, asiakasohjelman Java-tuen olemassaolo, Flash-tuen olemassaolo, näytön tarkkuus ja näytön värien lukumäärä. Nämä tiedot voidaan kerätä asiakaspäässä ajettavilla komentosarjakielillä. Asiakasohjelman ympäristömuuttujatietoja ei voida seurata palvelimelta, vaan niiden kerääminen vaatii käytön seurannan toteuttavan ohjelmakoo-

din ajamista käyttäjän asiakasohjelmassa. Asiakasohjelmalta saataviin ympäristömuuttujiin kuuluvat HTTP-otsikkotietojen mukana lähetettävät asiakasohjelma- ja käyttöjärjestelmätiedot. Tämä yksinkertaistaa monipuolisen käytön seurantajärjestelmän toteuttamista käyttäjän asiakasohjelmaan liittyen, koska kaikki tiedot saadaan samasta lähteestä. (Vrt. Karjalainen, 2010, ss. 83-84)

3.4 Tapahtumien seuranta

Tapahtumalla tarkoitetaan verkkopalvelun käyttöä, jolle voidaan antaa nimi. Tapahtuma voi ilmetä, kun käyttäjä hyödyntää jotain verkkopalvelussa olevaa toiminnallisuutta. Tapahtumalla voidaan tarkoittaa kahden verkkosivun välistä siirtymää jos tämä siirtymä on jollain tavalla merkittävä käytön seurantatiedollisesti, koska on mahdollista, että tapahtuma liittyy verkkopalvelussa olevaan tietosisältöön. Verkkopalveluissa olevia tietosisältöjä voidaan ainakin lisätä, muokata ja poistaa. Nämä kaikki voidaan mieltää omiksi tapahtumiksi, jotka liittyvät tähän kyseiseen tietosisältöön. Tapahtuma on mikä tahansa verkkopalvelussa tehtävä toimi, jota halutaan seurata.

Tapahtuman käsite liitetään usein verkkopalvelussa lähetettävään lomakkeeseen. Tapahtuma voi siis olla palveluun rekisteröityminen. Tällöin tapahtuma liittyy kiinteästi verkkopalvelun tiettyyn toiminnallisuuteen. Dynaamisissa verkkopalveluissa käytettävät teknologiat kuten AJAX mahdollistavat tapahtumien suorittamisen ilman, että sivuja joilla ne sijaitsevat ladataan uudelleen. Toimintoihin liittyvät tapahtumat luovat toisinaan verkkopalveluun uutta tietosisältöä kuten rekisteröitymisen tapauksessa, jolloin verkkopalveluun syntyy uusi käyttäjä. Rekisteröitymiseen liittyen, verkkopalveluun kirjautuminen on tapahtuma, joka liittyy verkkopalvelussa olevaan toimintoon, mutta ei lisää suoraan verkkopalvelussa olevan tietosisällön määrää. (Vrt. Alanto et al., 2010, s. 40)

3.4.1 Tapahtumasarjojen seuranta

Verkkopalvelun tapahtumia on mahdollista seurata sarjana peräkkäisiä tapahtumia. Tapahtumasarja voi olla esimerkiksi

- käyttäjän kulkema koko polku verkkopalvelussa (sisään \Rightarrow käyttö \Rightarrow poistuminen) tai
- verkkopalvelun käyttö ennen määriteltyä tapahtumaa (käyttö \Rightarrow tapahtuma)

Tapahtumasarjojen avulla pyritään selvittämään miten käyttäjät käyttävät verkkopalvelua. Tapahtumasarjojen avulla selvitetävät käyttöpolut saattavat esimerkiksi paljastaa verkkopalvelusta epäloogisia elementtejä. Tapahtumasarjojen avulla

voidaan havaita, että käyttäjän kulkema polku verkkopalvelussa tietyn verkkopalvelun ominaisuuden luokse ei ole se polku, joka sivustolla on tarkoitettu ensisijaiseksi poluksi ja jonka kautta toiminnallisuutta on ajateltu käytettävän.

Verkkopalveluja on mahdollista hyödyntää usein kahden roolin kautta, kirjautumattomana käyttäjänä ja kirjautuneena käyttäjänä. Tapahtumasarjoissa nämä kaksi roolia on mahdollista yhdistää ohjelmallisen istunnon (session) tai käyttäjän IP-osoitteen avulla. Tällöin voidaan vertailla käyttäjien toimia kirjautumattomina ja kirjautuneena ja havaita verkkopalvelusta tyypillisiä polkuja, jotka johtavat kirjautumistapahtumaan. Kirjautumattoman ja kirjautuneen käyttäjän yhdistäminen voi kuitenkin olla jossain tapauksissa vaikeaa. On mahdollista, että käyttäjän IP-osoite vaihtuu kesken verkkopalvelun käytön, joka on tavallista mobiiliverkoissa, kun verkkopalvelua käytetään mobiililaitteella. Tällöin käyttäjän yksilöiminen IP-osoitteen avulla vaikeutuu. Samasta IP-osoitteesta verkkopalvelua voi käyttää useampi henkilö, joten IP-osoite ei täysin yksilöi verkkopalvelun käyttäjää. Istuntoihin perustuva käyttäjän tunnistaminen on riippuvainen käyttäjän käyttämästä asiakasohjelmasta. Rinnakkaisesti käytettyjen asiakasohjelmien liittäminen yhteen ja samaan käyttäjään on vaikeaa. Toisaalta istuntojen elinikä on vain yhden selainistunnon mittainen, joten peräkkäisten käyttökertojen yhdistäminen toisiinsa on vaikeaa.

Tapahtumasarjoja on mielekästä seurata kvantitatiivisesta näkökulmasta eikä yksittäisiin käyttäjiin perustuen. On mahdollista seurata kahden verkkopalvelun näkymän välisten siirtymien määrää, eli kuinka monta kertaa sivulta “x” on siirrytty sivulle “y”. Kun tapahtumasarjoja seurataan puhtaasti kvantitatiivisesti, ei ole merkitystä kuka tapahtumasarjan suorittaa. Tapahtumasarjoja voidaan seurata useampien kuin kahden näkymän välisinä siirtyminä.

3.5 Yhteisöllisen käytön seuranta

Yhteisöllisen käytön seurannalla tarkoitetaan sellaista käytön seurantaa, joka tapahtuu verkkopalvelussa käyttäjien välillä. Yhteisöllisissä verkkopalveluissa käyttäjät vuorovaikuttavat toistensa kanssa. Yhteisöllisyys liittyy siihen, että verkkopalvelussa olevat käyttäjät luovat toisensa välille sosiaalisia kontakteja. Perinteisesti verkkopalveluissa vuorovaikutus tapahtuu verkkopalvelun käyttäjän ja verkkopalvelua ylläpitävän tahon välillä. Verkkopalvelujen yhteisölliset ominaisuudet ja verkkopalveluissa tapahtuvat käyttäjien väliset kontaktit luovat verkkopalvelun käytön seuraamiseen uusia näkökulmia. Yhteisöllisestä näkökulmasta on mahdollista seurata käyttäjien keskeisyyttä. Keskeisyys on yhteisöllisen käytön seurannan mittari, jolla kuvataan sitä miten usean muun käyttäjän kanssa verkkopalvelun yksittäinen käyttäjä on vuorovaikuttanut. (Silius et al., 2009, ss. 1-5)

Yhteisölliset verkkopalvelut kuten Last.fm³⁷, Facebook³⁸ ja IRC-Galleria³⁹ ovat kasvattaneet suosiotaan viime aikoina etenkin nuorten käyttäjien keskuudessa. Facebookin myötä sosiaalinen vuorovaikuttaminen on yleistynyt lisäksi keski-ikäisten ja jopa ikäihmisten keskuudessa. Tämän lisäksi erilaiset toimijat kuten yritykset, puolueet ja yhdistykset ovat löytäneet tiensä sosiaalisiin verkkopalveluihin. (Vrt. Silius et al., 2009, s. 1; vrt. Silius et al., 2010, s. 3) Sosiaalisuus ja vuorovaikuttaminen ovat Web 2.0:lle ominaisia piirteitä. Web 2.0 on yleinen määrittely, jota käytetään viitattaessa verkkopalveluissa oleviin yhteisöllisiin ominaisuuksiin. Yhteisöllisiä ominaisuuksia ovat verkkopalveluissa luotavat kaveruuskontaktit. (Esim. Graham, 2005; esim. Sharma, 2008; esim. O'Reilly, 2005; esim. Alanto et al., 2010, s. 15)

Toteutettaessa yhteisöllisen käytön seurantajärjestelmää on tärkeää, että seurantatiedosta pystytään erottamaan toisistaan tekijät ja toiminnot. Tekijöillä tarkoitetaan verkkopalvelun käyttäjiä, jotka toimivat verkkopalvelussa. Toiminto on jokin järjestelmän osa, jonka avulla tekijöiden on mahdollista vuorovaikuttaa keskenään. Toiminnot yhdistävät verkkopalvelun tekijöitä toisiinsa. Toiminto voi olla esimerkiksi keskusteluhuone, jossa on käynyt joukko käyttäjiä. Keskusteluhuoneessa käyneiden käyttäjien välille voidaan tällöin rakentaa löysä yhteyssuhde sen perusteella, että he ovat vierailleet samassa keskusteluhuoneessa, koska voidaan olettaa, että he jakavat tästä syystä jonkin yhteisen kiinnostuksen kohteen. (Silius et al., 2009, ss. 1-5)

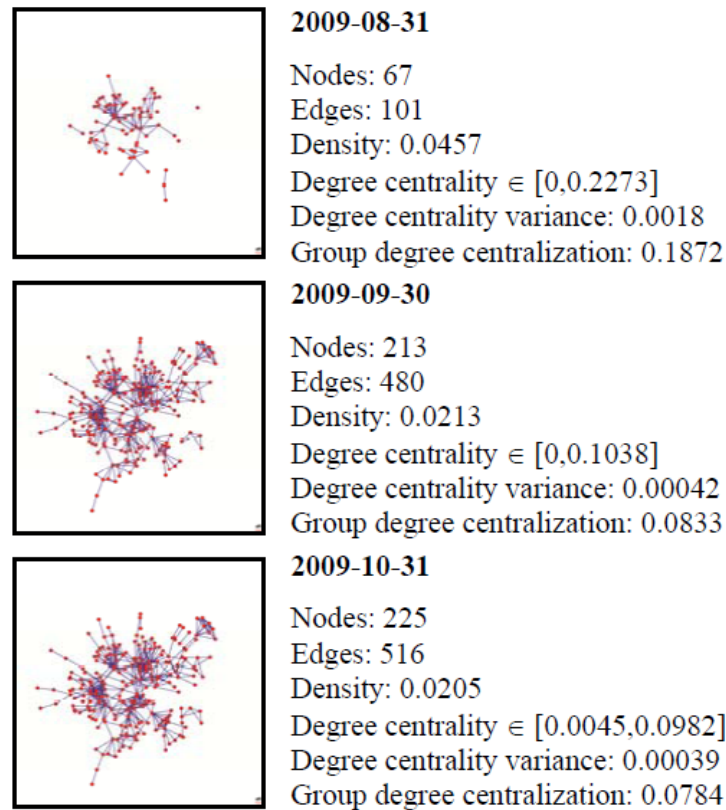
Yhteisöllisyyteen liittyvä käytön seuranta on hyvä tapa tehdä näkyväksi verkkopalvelun kasvua ja kehitystä. Kuvassa 3.3 nähdään miten Tampereen teknillisen yliopiston matematiikan laitoksella hypermedialaboratoriossa toteutettu ja opiskelijoille suunnattu yhteisöllinen TTY Piiri -verkkopalvelu⁴⁰ on kehittynyt kuukauden aikajaksoissa. Kuvassa 3.3 nähdään miten verkkopalvelun kehitystä voidaan kuvata verkkopalvelun yhteisöllisten ominaisuuksien avulla, kuten verkkopalvelun käyttäjien (kuvassa "Nodes") keskinäisten suhteiden avulla.

³⁷<http://last.fm/>

³⁸<http://facebook.com/>

³⁹<http://irc-galleria.net/>

⁴⁰<http://www.tut.fi/piiri/>



Kuva 3.3: Sosiaalisen vuorovaikutuksen seuranta verkkopalveluissa. (Silius et al., 2009, s. 5)

Tässä työssä ei keskitytä verkkopalvelun yhteisöllisten ominaisuuksien näkyväksi tekemiseen, koska tutkimuskohteena olevassa Erimenu.fi-verkkopalvelussa ei ole yhteisöllisyyteen kannustavia ominaisuuksia. Verkkopalvelujen yhteisöllisyyteen ja käyttäjien sosiaaliseen vuorovaikuttamiseen liittyvää tutkimustyötä on tehty enemmän edellä mainitun TTY Piiri -verkkopalvelun tapauksessa. (Silius et al., 2009, ss. 1-5)

3.5.1 Yhteisöllisyys verkkopalvelun käytön motivaattorina

Verkkopalvelun käytöstä saatavan tiedon avulla on mahdollista luoda palvelun käyttämiseen motivaatiota käyttäjien keskuudessa. Jos verkkopalvelun käyttäjät kokevat, että heidän toimintansa näkyy verkkopalvelussa, kasvaa usein heidän motivaationsa käyttää verkkopalvelua. Toisaalta kun käyttäjä näkee, miten muut ovat hyödyntäneet verkkopalvelua, saa käyttäjä tästä virikkeitä, joka kasvattaa hänen mielenkiintoaan käyttää palvelua. Mitä korkeampi motivaatio käyttäjällä on käyttää verkkopalvelua, sitä enemmän hän todennäköisesti käyttää sitä. (Vrt. Silius et al., 2010, ss. 3-4)

Verkkopalvelun seurantajärjestelmällä tätä käyttäjän motivaatiota käyttää verkkopalvelua voidaan kasvattaa tuomalla verkkopalvelussa olevaa yhteisöllisyyt-

tä näkyväksi. Jos käyttäjä kokee kanssakäyttäjien läsnäolon verkkopalvelussa, on hänen motivaationsa pysyä verkkopalvelussa mahdollisesti korkeampi. Seurantajärjestelmällä voidaan seurata verkkopalvelussa paikalla olevia käyttäjiä perustuen siihen, koska he ovat viimeksi ladanneet sivun verkkopalvelussa. Paikalla olevat käyttäjät on mahdollista näyttää verkkopalvelun käyttäjälle. Tämä ilmiö on vahvasti nähtävillä esimerkiksi Facebookissa, jossa muihin käyttäjiin liittyvää sisältöä päivitetään aktiivisesti käyttäjän näkyville. Verkkopalvelu tuo käyttäjän nähtävillä muihin käyttäjiin liittyvää sisältöä, joka lisää yhteisöllisyyden tuntemusta. Käyttäjälle kiinnostavien sisältöjen esittäminen ja tuominen käyttäjän saavutettaviksi vaatii kuitenkin monipuolista käytön seuranta järjestelmää. Verkkopalvelun täytyy tuntea käyttäjän ominaisuuksia ja tottumuksia, jotta se pystyy näyttämään verkkopalvelua käyttävälle käyttäjälle juuri sitä sisältöä verkkopalvelusta, joka on merkittävää kyseiselle käyttäjälle. (Vrt. Silius et al., 2010, ss. 3-4)

Erimenu.fi-verkkopalvelussa sosiaalisia ominaisuuksia esitellään palvelun etusivulla, jossa on nähtävillä käyttäjän ystäväpiiriin merkittyjen käyttäjien viime aikaisia tapahtumia. (Kuva 3.4)

	Lisätty:
 lisäsi avainsanan perheillal iselle ruokaohjeeseen Mausteinen pannacotta: Perusohje .	26 päivää sitten
lisäsi ruokaohjeen Mausteinen pannacotta: Perusohje omalle listalleen.	26 päivää sitten
lisäsi tuotteen TofuLine stracciatella 0,75 L omalle listalleen.	viime tammikuussa
lisäsi ruokaohjeen Riisilasagne merenherkuista: Perusohje omalle listalleen.	joulukuussa 2009

Kuva 3.4: Erimenu.fi:ssä sosiaalista käyttöä tuodaan käyttäjän näkyville avainsana ja oma lista -merkintöjen osalta. Käyttäjien yksityisyyteen liittyvät tiedot on sumennettu.

Verkkopalveluun toteutettavia käytön seurantaominaisuuksia voidaan tuoda käyttäjien saataville ja niiden avulla voidaan luoda tai jopa ohjata verkkopalvelun käyttöä. Kuvassa 3.4 nähdään miten verkkopalvelun muiden käyttäjien toimet esitetään verkkopalveluun kirjautuneelle käyttäjälle. Tämä saattaa ohjata käyttäjää niin, että hän ei enää lisää itse avainsanoja, koska ei halua tulla palvelussa nähdyksi. Toisaalta on mahdollista, että käyttäjä ryhtyy lisäämään itse enemmän avainsanoja, koska kokee, että tästä on hyötyä muille käyttäjille. Aliluvussa 3.5 esiin tuodussa TTY Piiri -verkkopalvelussa yhteisöllistä käytön seurantatietoa tuotiin käyttäjien saataville siten, että luotujen käyttäjäsuhteiden avulla järjestelmä pyrki ehdottamaan käyttäjälle muita käyttäjän mahdollisesti tuntemia käyttäjiä. Kun järjestelmä helpottaa käyttäjän toimintaa, luo tämä verkkopalvelusta yhteisöllisesti monipuolisemman ja tekee siten käytön seurantatiedosta rikkaampaa.

Yhteisöllisyyden esille tuomisella voi olla negatiivisia vaikutuksia verkkopalvelun käytölle. Esimerkiksi jos verkkopalveluun tuotetaan sellainen ominaisuus tai toiminto, joka tuo käyttäjien toimintaa esille sellaisella tavalla, joka ei ole käyttäjien mieleen, saattaa tämä johtaa siihen, että verkkopalvelun käyttäjät eivät enää käytä tätä ominaisuutta. Tämänkaltainen toiminnallisuus saattaisi olla esimerkiksi

IRC-Galleriasta⁴¹ löytyvä toiminto, joka kertoo käyttäjän profiilissa vierailleet käyttäjät. On mahdollista, että käyttäjät eivät halua, että tietty henkilö, jonka profiilissa he vierailevat, saa tietää siitä, että he ovat käyneet katsomassa hänen profiiliaan. Tämä saattaa johtaa siihen, että profileja selataan tämän johdosta enemmän palveluun kirjautumattomana, joka vaikeuttaa käyttäjien yksilöintiä. Verkkopalvelun käyttöä näkyvämmäksi tekeviä ominaisuuksia toteutettaessa on arvioitava mahdolliset negatiiviset tekijät. (Vrt. Silius et al., 2010, ss. 3-4)

3.6 Seurantatiedon tallennus

Tiedon tallennusmuodon määrittely on tärkeä palanen käytön seurannan toteutuksessa. Tallennusmuodolla tarkoitetaan, mihin ja minkälaisella merkkauksella kerättävät tiedot tallennetaan. Aliluvussa 2.1 mainittiin, että tiedon tallennukselle on olemassa määriteltyjä muotoja kuten ELF (Extended Log Format) ja CLF (Common Log Format). Nämä ovat toimivia tiedon tallennusmuotoja, jos halutaan kerätä käytön seurantatietoa palvelin tasolla.

Monipuolisempien ja monimutkaisempien sovellustason käytön seurantatietojen tallennukseen ei ole olemassa kattavaa mallia, koska sovellukset ovat keskenään hyvin erilaisia toiminnoiltaan sekä toteutukseltaan. Esimerkiksi verkkopalveluun, jossa ei käytetä hyväksi SQL-pohjaista tietokantaa, ei ole perusteltua liittää tätä teknologiaa pelkästään käytön seurantajärjestelmää varten. Tiedon tallennukseen ei voida siis määritellä yksiselitteisesti toimivaa tallennusmuotoa, jota voitaisiin hyödyntää yleisesti.

Tiedon tallennusmuodolle voidaan määritellä hyviä periaatteita. On tärkeää, että tallennetut tiedot ovat helposti hallittavissa ja niiden käsittely on nopeaa, koska kerättyä tietoa halutaan usein visualisoida ja käsitellä erilaisilla työkaluilla. Nämä työkalut vaativat, että niille annettu tieto on tiettyssä muodossa, jolloin kerättyä tietoa täytyy jäsentää niille sopiviksi. Tallennusmuodon ennakointi on vaikeaa, koska ei tiedetä vielä tarkkaan millä kaikilla välineillä tietoa tullaan käsittelemään ja yleensäkin mitä kaikkea tietoa saadaan kerättyä.

3.6.1 Tietokantatallennus

Useimmat verkkopalvelut käyttävät hyväkseen SQL-pohjaista tietokantaa. Tietokantaan tallennetaan verkkopalvelun tietosisällöt, josta niitä tarvittaessa haetaan. Tietokantaa voidaan käyttää lokitiedon tallennukseen. Tietokannalle on ominaista, että tietokannassa on tauluja, jotka sisältävät rivejä ja sarakkeita. Yksittäiset sivulataukset voidaan tallentaa tietokantariveiksi. Tällöin yksittäinen rivi sisältää tiedot yksittäisestä sivulatauksesta ja rivin sarakkeet kertovat tähän sivulataukseen

⁴¹<http://irc-galleria.net/>

liittyvät tiedot. Tiedot voidaan tallentaa sarakkeittain niin, että yksittäisessä sarakkeessa on yksittäinen tieto. Toinen vaihtoehto on tallentaa sivulataukseen liittyvät tiedot, jossain muodossa sarjallistettuina yhteen sarakkeeseen. Jälkimmäinen tapa kasvattaa jälkiprosessoinnin määrää kun sarakkeessa oleva tieto täytyy jäsentää. Ensin mainittu keino saattaa kasvattaa sarakkeiden määrän hyvin suureksi, jolloin tietokantaan tehtävien kyselyiden kesto kasvaa ajallisesti.

3.6.2 Tiedostotallennus

Verkkopalvelun käytön seurantatieto voidaan tallentaa levyllä oleviin tiedostoihin. Näin toimivat esimerkiksi HTTP-palvelinohjelmistot. Tiedostojen etu on, että suoraan levyllä yksittäisestä tiedostosta lukeminen on nopeampaa kuin tietokannasta, jossa tiedon lukeminen tapahtuu erillisen tietokantamoottorin avulla. Haittapuolena on tiedon alhainen rakenteisuus ja käsiteltävyys, jonka johdosta luetun tiedon prosessointi voi olla hyvin vaikeaa ja hidasta. Kirjoitettaessa seurantatietoa tiedostoon toimitaan yleensä niin, että yksittäisten tapahtumien erottimena käytetään rivinvaihtoa.

3.7 Käytön seurantajärjestelmän käyttöönotto

Verkkopalvelun käytön seuraaminen on yleensä koko järjestelmää koskeva ominaisuus. Näin on koska verkkopalvelun käytöstä tallennetaan usein tieto palveluun liittyvistä sivulatauksista, joka tarkoittaa sitä, että käytön seurantajärjestelmä on läsnä aina kun verkkopalvelua hyödynnetään. Käytön seurannan käyttöönottaminen vaikuttaa siis kaikkiin verkkopalvelun osiin, joten on tärkeää, että toteutettu järjestelmä testataan huolellisesti ennen käyttöönottoa. On mahdollista, että toteutettu järjestelmä rikkoo jonkin olemassa olevan toiminnallisuuden verkkopalvelussa integroituaan siihen. Käytön seurantajärjestelmän käyttöönotossa on hyvä huomioda, että kaikkia verkkopalvelun toimintoja ja osia ei välttämättä ole tarpeellista seurata. Seurantajärjestelmä on mahdollista ottaa käyttöön vain verkkopalvelun niissä osissa, joista seurantatietoa halutaan kerätä. Kun käytön seurantaa ei sisällytetä niihin verkkopalvelun osiin, joissa se on tarpeeton vähennetään riskiä siihen, että käyttöönotettava seurantajärjestelmä rikkoo verkkopalvelussa olevia toimintoja. (Vrt. Valtionvarainministeriö, 2007, s. 82; vrt. Croll & Sean, 2009, s. 5-6)

3.7.1 Seurantajärjestelmän toteuttaminen uuteen verkkopalveluun

Käytön seurantajärjestelmän toteuttaminen verkkopalveluun kehitystyön ohessa on oletusarvoisesti helpompaa kuin verkkopalvelun elinkaaren myöhemmissä vaiheissa. Tästä johtuen järjestelmän toiminnot voidaan lähtökohtaisesti toteuttaa tukemaan niiden seuraamista. Yleisesti toteutettaessa järjestelmän toiminnallisuuksia

on tärkeää, että niiden käytöstä tallennetaan niitä käyttävän käyttäjän tiedot ja toiminnallisuuden käyttö- tai luontiaikatiedot. Sama periaate pätee verkkopalvelussa oleviin tietosisältöihin. Tietosisältöjen yhteyteen tulee tallentaa tieto sisällön luoneesta käyttäjästä ja sisällön luontiajasta. Jos verkkopalvelussa olevia tietosisältöjä on mahdollista muokata, tulee tiedon muokkaaja ja muokkaus aika tallentaa. Tiedon lisääjä ja lisäämisaika eivät välttämättä ole verkkopalvelun kannalta tärkeitä tietoja, mutta tietojen tallentaminen on tarpeellista aina kun verkkopalvelun käyttöä halutaan seurata. Huomioitavaa on, että tietosisältöihin liittyvien lisätietojen lisääminen tietosisältöihin voi olla jälkikäteen mahdotonta, joten niiden tallentaminen mahdollista myöhempää tarvetta silmällä pitäen on kannattavaa.

3.7.2 Seurantajärjestelmän sisällyttäminen verkkopalveluun

Verkkopalvelun käytön seuranta ei nähdä verkkopalvelun pääasiallisena toimeina, koska seurantaan liittyviä kysymyksiä pohditaan verkkopalvelujen toteutuksessa vasta, kun verkkopalvelun muut toiminnot on saatu toteutettua. Seurantajärjestelmän kohdalla puhutaan teknisesti ominaisuudesta, joka on läsnä kaikkia verkkopalvelun toimintoja käytettäessä, joten sen käyttöönotto olemassa olevaan verkkopalveluun vaatii verkkopalvelun ominaisuuksien kattavan testauksen.

Käyttäjälähtöisesti tarkasteltuna käytön seurantajärjestelmä on verkkopalvelun ominaisuus, johon käyttäjä ei ole suostunut hyväksyessään verkkopalvelun käyttöehdot. Tällöin verkkopalvelun käyttäjille tulee tiedottaa verkkopalvelun käytön seurantajärjestelmän käyttöönotosta (katso aliluku 2.8). Käytön seurantajärjestelmä ei saa vaikuttaa verkkopalvelun käyttöön, koska kyse ei ole verkkopalvelun pääasiallisesta toiminnallisuudesta. Jos verkkopalvelun käytön seurantajärjestelmä hidastaa verkkopalvelun käyttöä merkittävästi, huononee verkkopalvelua käyttävän käyttäjän käyttökokemus.

3.8 Käytön seurantatiedon vienti ja tuonti

Käytön seurantatiedon viennillä ja tuonnilla tarkoitetaan käytön seurantatiedon siirtämistä käytön seurantajärjestelmän ja muiden järjestelmien välillä. Käytön seurantatietoa on mahdollista viedä seurattavasta järjestelmästä muihin järjestelmiin käsiteltäväksi ja käytön seurantatietoa on mahdollista tuoda käytön seurantajärjestelmään. Käytön seurantatiedon siirtäminen voi olla tarpeellista, kun järjestelmän keräämälle tiedon muodolle täytyy tehdä jokin muunnos. Tällöin käytön seurantatieto viedään ensin järjestelmästä muunnoksen tekevään järjestelmään, jonka jälkeen muunnettu käytön seurantatieto tuodaan takaisin järjestelmään. Käytön seurantatiedon siirtämisessä järjestelmien välillä on tärkeää, että seurantatieto pystytään esittämään rakenteellisessa muodossa, koska tällöin sen käsittely helpottuu.

XML (eXtensible Markup Language) on koneellisesti tuotettava ja luettava kieli. Kielen merkkauksessa on otettu huomioon luettavuus ihmisen näkökulmasta. XML-dokumentti koostuu merkkauksesta ja merkkidatasta. Alla esimerkki yksinkertaisesta XML-merkkauksesta, josta nähdään, että XML-merkaus koostuu juurielementistä, juurielementin alla olevista lapsielementeistä, elementtien arvoista ja elementtien attribuuteista. XML:ssä elementti merkitään kulmasulkeiden avulla. Attribuutit tulevat XML-elementin sisään elementin nimen jälkeen. (Fallside & Walmsley, 2004)

```
<?xml version="1.0" encoding="utf-8"?>          <!-- XML-julistus -->
<logdatas>                                       <!-- juurielementti -->
  <logdata id="id01" date="2009-06-28">         <!-- lapsielementti -->
    Sivulataustapahtuma                          <!-- jossa attribuutteja -->
  </logdata>                                       <!-- ja niiden arvoja -->
</logdatas>
```

XML soveltuu lokitiedon esittämiseen ja siirtämiseen järjestelmien välillä, koska se sisältää riittävän kattavat välineet tiedon esittämiseen. XML on yleisyytensä vuoksi helposti käsiteltävissä eri ohjelmilla ja merkkaukseltaan tarpeeksi kevyt soveltuakseen suurehkojen tietomäärien käsittelemiseen (vrt. Everi, 2006, ss. 13-14). Käytön seurantatiedon vientimuotoa määriteltäessä on tärkeää kuitenkin miettiä, missä vietyä tietoa tullaan käyttämään. Kun käytön seurantatieto viedään järjestelmästä pääsääntöisesti taulukkolaskentaohjelman käsiteltäväksi, on hyvä harkita CSV-tiedostomuodon (Comma-Separated Values) hyödyntämistä tiedon vientimuotona, koska CSV-tiedostomuodossa oleva lokitieto voidaan tuoda sellaisenaan useimpiin taulukkolaskentaohjelmiin (Shafranovich, 2005, s. 2).

Käytön seurantatiedon käsittelemiselle on ominaista tiedon suuri määrä. Käytön seurantatieto sisältää tiedon järjestelmään kohdistuneista sivulatauksista, jolloin jos järjestelmällä on paljon aktiivisia käyttäjiä, kasvaa käytön seurantatiedon määrä hyvin nopeasti ja hyvin suureksi. Jos esimerkiksi käsitellään verkkopalvelua, jolla on 100 päivittäistä kävijää, jotka keskimäärin lataavat 10 sivua päivässä kuukauden ajan päädytään seuraavaan lopputulokseen:

$$k = 100 \cdot 10 \cdot 31 = 31\,000.$$

Tietokantaan tallennetaan kuukauden aikana mainituilla käyttömäärillä 31 000 tietokantariviä. Käytön seurantatiedon määrän suuruus aiheuttaa ongelmia esimerkiksi kun tieto halutaan koota yhteen ja viedä järjestelmästä tiedostona. Käsiteltäessä näin suuria määriä tietoa on tiedon esitystavan valinnalla merkitystä.

3.8.1 Seurantatiedon vienti järjestelmästä

Käytön seurantatiedon viennillä tarkoitetaan, että käytön seurantajärjestelmän keräämä tieto otetaan järjestelmästä ulos. Seurantatieto voidaan koostaa käytön

seurantajärjestelmästä tiedostoon, joka tarjotaan järjestelmän kautta riittävän käyttöoikeudet omaavalle käyttäjälle ladattavaksi. Seurantatiedon vientiä voidaan hyödyntää, kun käytön seurantatietoa halutaan visualisoida jossain verkkopalvelun ulkopuolisessa järjestelmässä. Seurantatiedon vientimuotoa määriteltäessä on hyvä tietää mihin ja millä tavalla järjestelmästä vietyä tietoa käytetään. Kun halutaan viedä verkkopalvelusta luodut ystävyysuhteet seurantatiedon yhteyteen, on oleellista pystyä yhdistämään tiedot käyttäjistä, jotka liittyvät luotuihin ystävyysuhteisiin. Käyttäjätietoihin kuuluvat tavallisesti käyttäjän demografiset tiedot kuten asuinkunta, sukupuoli ja ikä. (Vrt. Silius et al., 2009, ss. 5-6)

Lokitiedon vientiin on mahdollista toteuttaa oma toiminnallisuutensa suoraan verkkopalvelun yhteyteen, jossa seurantatiedon keräys tapahtuu. Tällöin toiminnallisuus on helposti sitä tarvitsevien käyttäjien saavutettavissa. Vaihtoehtoisesti käytön seurantatiedon viennissä voidaan hyödyntää verkkopalvelun tietokannanhallintajärjestelmän ominaisuuksia. Yksinkertaisimmillaan käytön seurantatiedon vienti voidaan toteuttaa hyödyntäen tietokannanhallintajärjestelmiin sisäänrakennettuja tiedon vientitoiminnallisuuksia. Järjestelmän käyttäessä MySQL⁴²-tietokannanhallintajärjestelmää, on vienti mahdollista suorittaa graafisessa käyttöliittymässä MySQL:ään integroituvan phpMyAdmin⁴³-ohjelmiston avulla. PhpMyAdmin tarjoaa laajat ominaisuudet tietosisältöjen vientiin järjestelmästä eri muodoissa, joten sen hyödyntäminen sellaisenaan käytön seurantatiedon vientiin on usein riittävää. Tietokantaviennin heikkous on se, että se perustuu hyvin pitkälle tietokannan malliin eli taulurakenteeseen. Tällä tarkoitetaan, että haluttaessa viedä sellaista tietoa, joka on tallennettu tietokannassa useampaan kuin yhteen tauluun joudutaan joko kirjoittamaan tietokantakysely, joka palauttaa tämän tuloksen tai jälkikäsittelemään useita tulosjoukkoja. Onko käytön seurantatiedon vienti mielekästä mahdollistaa ulkopuolisessa järjestelmässä vai suoraan kerättävänä olevan järjestelmän yhteyteen, on tapauskohtaista ja perustuu paljon siihen kuka ja miten usein käytön seurantatiedosta koottuja visualisointeja hyödynnetään. Jos tarkoitus on, että käytön seurantatiedosta toteutettavat visualisoinnit tulevat verkkopalvelun käyttäjien hyödynnettäviksi, on mielekästä, että ne ovat saatavilla verkkopalvelusta. Toisaalta jos käytön seurantatiedolla halutaan osoittaa yrityksen johdolle, että järjestelmään tehty muutos näkyy järjestelmässä tietyllä tavalla, ei seurantatiedon visualisointien sisällyttäminen järjestelmän yhteyteen välttämättä ole yhtä tarpeellista.

3.8.2 Seurantatiedon tuonti järjestelmään

Käytön seurantatiedon tuonti järjestelmään on harvinaisempaa kuin käytön seurantatiedon vienti järjestelmästä. Kerätyn tiedon tuonti saattaa kuitenkin olla

⁴²<http://www.mysql.com/>

⁴³<http://www.phpmyadmin.net/>

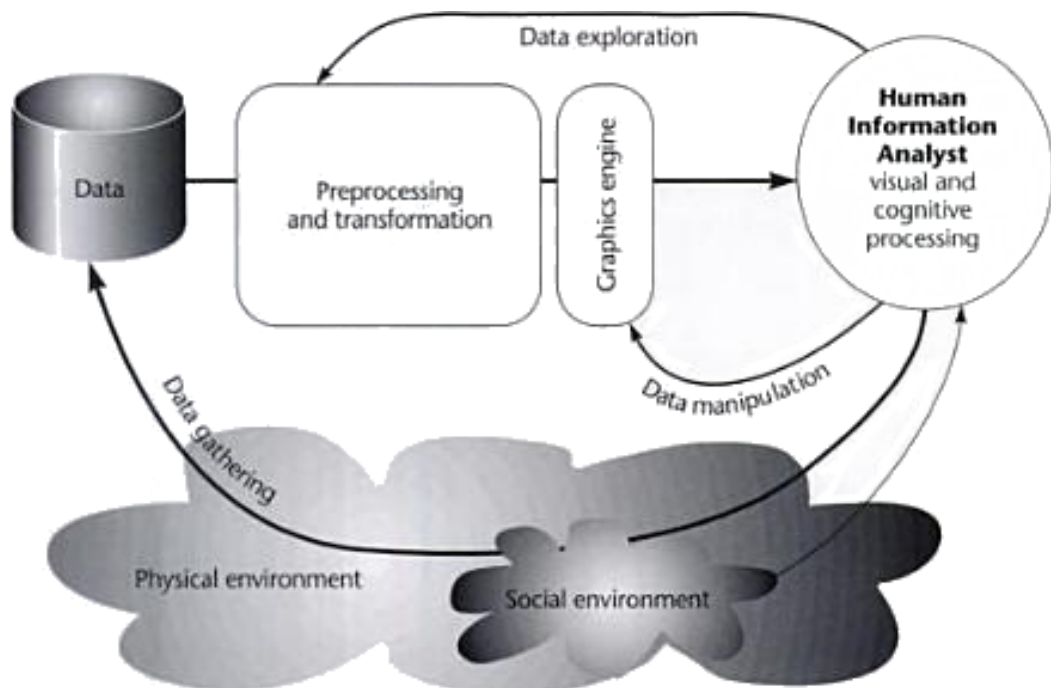
tarpeellista. Kun verkkopalveluun ollaan toteuttamassa uutta seurantajärjestelmää, vanhan järjestelmän keräämät tiedot täytyy tuoda uuteen järjestelmään. Toisaalta taas kerättyä käytön seurantatietoa täytyy järjestelmän päivityksestä johtuen muuntaa toiseen muotoon. Muunnos toteutetaan itse verkkopalvelun ja käytön seurantajärjestelmän ulkopuolella, koska niistä ei löydy muuntamiseen tarvittavia työkaluja. Tässä tapauksessa seurantatieto viedään ensin järjestelmästä ja tuodaan sitten muunnettuna takaisin järjestelmään. Kun verkkopalvelussa käytetään tiedon tallennukseen tietokantaa, on seurantatiedon tuonti mahdollista toteuttaa tietokantakyselyiden avulla. Käytön seurantatiedon tuontiin on harvoin tarvetta toteuttaa erillisiä järjestelmiä esimerkiksi verkkopalvelun yhteyteen.

4 Verkkopalvelun käytön visualisointi

Käytön seurantatiedon visualisoinnilla tarkoitetaan kerätyn käytön seurantatiedon koostamista ja näyttämistä helposti ymmärrettävässä muodossa. Seurantatiedon visualisointi on tarpeellista haluttaessa kuvata verkkopalvelun käyttöä käytön seurantatiedon avulla. Kerätty tieto pitää sisällään vain yksittäisiä tietoalkioita, joiden perusteella verkkopalvelun käytön tulkinta on vaikeaa tai mahdotonta. Visualisoinnin on tarkoitus havainnollistaa ja kuvata käytön seurantatiedossa oleva verkkopalvelun käyttötieto, jotta sen tulkitseminen ja analysointi on mahdollista. Visualisointi voi olla käytön seurantatiedosta piirretty kuvaaja tai laskettu tunnusluku.

4.1 Käytön seurantatiedon keräysprosessi

Käytön seurantajärjestelmän kerääminen ei rajoitu vain käytön seurantatiedon keräämiseen vaan se voidaan nähdä kokonaisvaltaisena prosessina, joka käsittää erilaisia vaiheita ja elementtejä. Kerättävä tietojoukko toimii vain yhtenä tekijänä tässä prosessissa. Prosessiin kuuluvat verkkopalvelun käyttöympäristö (physical and social environment), käytön seurantatieto (data), tiedon prosessointi (preprocessing and transformation), tiedon visualisointi (graphics engine) ja tiedon analysointi (human information analyst). Prosessin vaiheet ja niiden väliset yhteydet esitetään kuvassa 4.1.



Kuva 4.1: Visualisointilähtöinen malli tietojoukon esittämiseen. (Ware, 2004, s. 12)

Visualisointitieteissä visualisointiprosessi kuvataan neljän perusvaiheen kautta. Nämä neljä vaihetta ovat

- tiedon kerääminen ja tiedon tallentaminen
- esiprosessointi ja tiedon muuntaminen ymmärrettävään muotoon
- tietojoukon graafinen käsittely ja
- visualisoidun tiedon analysointi. (Ware, 2004, ss. 4-12)

Toteutettaessa verkkopalvelun käytön seurantajärjestelmää edellä kuvattu prosessimalli on käyttökelpoinen, koska verkkopalvelujen käytön seurantajärjestelmän toteutusprosessi noudattaa hyvin paljon samoja periaatteita. Verkkopalveluissa toimitaan jossain käyttöympäristössä, jossa luodaan kerättävää tietoa, jonka jälkeen kerättävä tieto prosessoidaan sopivaan muotoon ja syötetään halutulle visualisointivälineelle, joka prosessoi siitä ihmiselle analysoitavan muodon. Kuvassa 4.1 nähtävät visualisointiprosessin mukaiset takaisinkytkennät vaiheiden välillä pitävät paikkansa verkkopalvelujen käytön seurantajärjestelmässä. Kerätyn tiedon manipulointi visualisointien perusteella ja uudelleen visualisointi on tavallinen toimenpide verkkopalvelujen käytön visualisoinnissa, koska kerätylle tiedolle on helpompaa etsiä kokeilemalla sopiva visualisointi kuin määrittelemällä se kerätyn tiedon perusteella.

4.2 Visualisointi

Visualisoinnilla tarkoitetaan sellaista tiedon esitystapaa, joka on ihmisen näköaistille luettavassa muodossa. Yleisimmin visualisoinnin voidaan mieltää tarkoittavan minkä tahansa tiedon esittämistä sellaisessa muodossa, jossa se on aikaisempaa ymmärrettävämmässä muodossa.

Visualisointikeinoja on monia erilaisia. Visualisointi voi olla esimerkiksi

- tulevan päivän säätä visualisoiva sääkartta
- laskennan tuloksia visualisoiva taulukkolaskentakaavio tai
- käsitteiden välisiä suhteita visualisoiva käsitekartta.

Visualisointi ja visualisoinnit ovat hyvin lähellä ihmisen arkipäiväisiä toimia ja niitä käsitellään huomaamatta päivittäin. Hyvän visualisoinnin tunnusmerkki on, ettei sen olemassaoloa huomaa. Visualisoitaessa monimutkaisia ilmiöitä, joilla ei ole arkipäiväistä merkitystä, tähän päästään harvoin.

Tufte (2001) on määritellyt hyvän visualisoinnin ominaisuuksiksi seuraavat tekijät:

- hyvä tieto/muste -suhde
- mahdollisimman vähäinen kaavioteknisyys
- esteettisyys
- ei toistoa ja
- monitoiminnallisten elementtien käyttäminen.

Visualisoinnilla pyritään lisäämään ymmärrystä visualisoinnin kohteena olevasta ilmiöstä. On kuitenkin mahdollista, että visualisoinnilla johdetaan harhaan vääristelemällä visualisointia sopivalla tavalla. Esimerkiksi kuvaajien kohdalla tätä voidaan tehdä muuntelemalla asteikkoa visualisointien kesken. Visualisointi peittää alleen yksittäisten tietoalkioiden arvot, joten näiden vääristely tai huomiotta jättäminen on visualisoinnin avulla mahdollista.

Verkkopalvelun käytön seurantatiedon visualisoinnissa yksittäiset tietoalkiot kootaan loogiseksi kokonaisuuksiksi ja esitetään käyttäjälle erilaisia tunnuslukuina tai kuvaajina. (Siirtola, 2007, s. 5)

4.2.1 Tiedon visualisointi ja tiedon analysointi

Käytön seurantatietoa käsiteltäessä on hyvä tehdä ero mitä tarkoitetaan käytön seurantatiedon visualisoinnilla ja toisaalta mitä tarkoitetaan käytön seurantatiedon analysoinnilla. Tiedon visualisointi tarkoittaa, että verkkopalvelun käytöstä kerätty seurantatieto koostetaan helpommin luettavissa olevaan muotoon. Käytön seurantatiedon analysointi tarkoittaa, että seurantatiedosta tuotettuja visualisointeja luetaan ja pyritään niistä saatavalla tiedolla tekemään perusteltuja johtopäätöksiä. Ero käsitteiden välillä on hyvä tehdä, koska käytön seurantatietoa käsittelevässä kirjallisuudessa paneudutaan usein juuri tiedon analysointiin. Luvun 4 tarkoituksena on kertoa seurantatiedon visualisoinnista. Kirjallisuudessa käytön seurantaa käsitellään analysoinnin näkökulmasta, koska tämä on yritysmaailman lähtökohdista kiinnostavampaa. (Vrt. Kaushik, 2010, ss. 17-19)

4.2.2 Visualisointitekniikat Internetissä

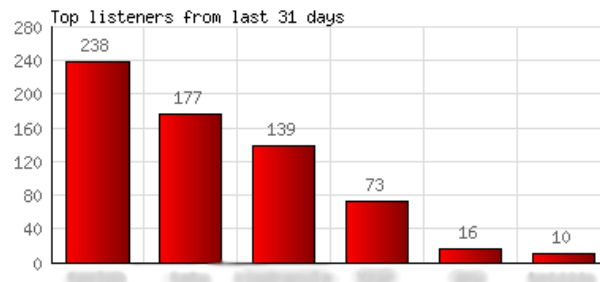
Internetissä käytön seurantatiedon dynaamisessa visualisoinnissa käytetään pääasiassa kahta teknologiaa: JavaScript ja Adobe Flash. Nämä teknologiat mahdollistavat dynaamisten ja interaktiivisten sisältöjen sisällyttämisen WWW-sivuille.

Molemmilla teknologioilla on vahvuutensa, mutta niiden käyttö saattaa toisaalta vaatia erillisten lisäosien asentamista asiakasohjelmiin. Adobe Flashiä hyödynnetään etenkin silloin, kun halutaan luoda erilaisia animaatiota. JavaScriptiä taas käytetään nykyään usein visualisointiin erilaisten kirjastojen tukemana. Tällä hetkellä jQuery⁴⁴, jQuery UI⁴⁵ ja ProtoType⁴⁶ -nimiset JavaScript-kirjastot tarjoavat suoraan tai välillisesti monipuolisia mahdollisuuksia tietojoukkojen käsittelyyn. (Vrt. Alanto et al., 2010, s. 40; *Ajax Widget Frameworks Survey*, 2009)

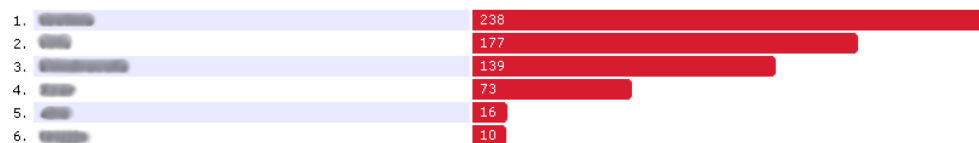
JavaScriptin ja Adobe Flashin lisäksi verkkopalvelun käyttöä on mahdollista visualisoida staattisilla kuvilla. Useimmissa ohjelmointikielissä on mahdollista tuottaa .jpg tai .png kuvasisältöjä, joihin kerätty seuranta-tieto voidaan sisällyttää.

4.2.3 Kuvaajat verkkopalvelujen visualisoinnissa

Puhuttaessa käytön seuranta-tiedon visualisoinnista mielletään visualisointi usein tietoa esittäväksi kuvaajaksi. Käytön seuranta-tiedon visualisoinnissa voidaan hyödyntää erilaisia kuvaajatyyppejä. Tavallisesti käytön seuranta-tietoa voidaan visualisoida viivadiagrammien, pylväsdiagrammien, suunnattujen graafien ja piirakka-diagrammien avulla. Kuvaajatyypin valinnalla on merkitystä siihen miten kuvaajaa tulkitaan. Kuvissa 4.2 ja 4.3 nähdään miten sama tieto voidaan eri tavalla esittämällä saada näyttämään hyvin erilaiselta.



Kuva 4.2: Esimerkki pystyyn asetellusta pylväsdiagrammista. Käyttäjien yksityisyyteen liittyvät tiedot on sumennettu.



Kuva 4.3: Esimerkki vaakatasossa olevasta pylväsdiagrammista. Visualisoinnissa pienimmät arvot ovat kuvaajateknisistä syistä skaalattu samaan leveyteen. Käyttäjien yksityisyyteen liittyvät tiedot on sumennettu.

⁴⁴<http://jquery.com/>

⁴⁵<http://jqueryui.com/>

⁴⁶<http://www.prototypejs.org/>

Kuvaajien avulla voidaan vaikuttaa siihen, millä tavoin tieto ymmärretään ja miten helposti se on omaksuttavissa. Kuvaajatyypin valinnalla on vaikutusta, miten visualisointeja voidaan vertailla keskenään. Usein eri tyyppisiin kuvaajiin piirrettyjä tietojoukkoja on vaikeaa vertailla keskenään, kun taas samankaltaisiin kuvaajatyyppeihin sijoitettuja tietoja on helpompaa vertailla toisiinsa. Kuvaajatyyppejä tulee valita esittävän tiedon mukaan ja esimerkiksi piirakkadiagrammit soveltuvat sellaiseen visualisointiin, jonka avulla halutaan vertailla tietojoukon alkiodien määrien välisiä osuuksia.

4.2.4 Tunnusluvut verkkopalvelujen visualisoinnissa

Käytön seurantatiedosta on mahdollista laskea erilaisia tunnuslukuja. Nämä tunnusluvut voidaan nähdä omanlaisinaan visualisointeina, koska ne koostavat kerättyä seurantatietoa ja kasvattavat ymmärrystä verkkopalvelun kokonaistilasta. Erilaisia tunnuslukuja voidaan määritellä verkkopalvelukohtaisesti. Yleisiä tunnuslukuja, jotka voidaan laskea useimmille verkkopalvelun tietosisällöille, ovat tietojoukon keskiarvo, mediaani, suurin arvo, pienin arvo, ensimmäinen arvo ja viimeisin arvo.

Keskiarvo ja mediaani kuvaavat tyypillistä arvoa tietojoukossa. Keskiarvo ja mediaani voidaan laskea verkkopalveluun kuukausittain rekisteröityneiden käyttäjien määristä. Tietojoukon suurin ja pienin arvo kuvaavat sitä miten paljon variaatiota tietojoukossa on. Suurimman ja pienimmän arvon avulla voidaan tutkia minkä kuukauden aikana verkkopalveluun on rekisteröitynyt eniten ja vähiten käyttäjiä. Tietojoukon ensimmäisen ja viimeisen arvon avulla voidaan havainnollistaa, miten verkkopalvelussa olevaa toiminnallisuutta on käytetty ajan mukaan.

4.2.5 Käyttäjien segmentointi

Segmentointi on markkinointiin liittyvä perustermi. Segmentoinnilla tarkoitetaan asiakkaiden jakamista ryhmiin heidän ominaisuuksiensa mukaan. Segmentointi voi perustua koviin ja pehmeisiin ominaisuuksiin. Kovat segmentointiominaisuudet ovat pysyvämpiä kuin pehmeät ominaisuudet ja niihin kuuluvat käyttäjän ikä ja sukupuoli. Pehmeisiin ominaisuuksiin lasketaan kuuluvan harrastukset ja mielipiteet. Verkkopalveluissa segmentointia voidaan tehdä hyvin monella tavalla. Useimmiten segmentointia käytetään verkkopalvelun käyttäjien jakamiseen erilaisiin käyttäjäryhmiin heidän demografisten piirteittensä perusteella. Segmentoinnin apuna käytettäviä käyttäjien demografisia piirteitä ovat mm: ikä, sukupuoli ja asuinpaikka. Segmentointia voidaan kuitenkin harrastaa käytön seurantatiedon visualisoinnissa muillakin tavoilla kuin käyttäjän demografisten tietojen kautta. Kun käyttäjä nähdään käytön seurannassa asiakasohjelmana, demografisiin tietoihin voidaan lisätä muun muassa käyttöjärjestelmä, selain ja näytön resoluutio, jolloin käyttäjät voidaan jaotella

näiden tietojen mukaan. (Vrt. Dickson & Ginter, 1987, ss. 1-10)

Tärkeää segmentoinnissa on valita käytetyt segmentit järkevällä tavalla, koska niiden valinnalla on oleellinen merkitys visualisoitavan tietojoukon ulkonäköön. Kun halutaan kuvata lapsille suunnatun verkkopalvelun ikäjakaumaa ja segmenteiksi valitaan 0-20 vuotiaat, 20-40 vuotiaat, 40-60 vuotiaat, voidaan visualisoinnin perusteella helposti todeta ettei tietojoukon segmentointi ole onnistunut, koska pääosin kaikki palvelun käyttäjät osuvat yhteen segmenttiin. Esitetyt segmentit saattavat kuitenkin toimia jonkin toisen verkkopalvelun käyttäjien ikäjakauman kuvaamiseen. Segmentointi on tietolähtöistä. (Vrt. Dickson & Ginter, 1987, ss. 1-10)

Segmentointi on ihmiselle luonteva tapa ymmärtää maailmaa. Ihmiselle on luontevaa koostaa asioista suurempia helpommin hallittavia kokonaisuuksia. Toisinaan käsiteltävinä oleva tietojoukko segmentoituu luontevasti. Resoluutio on luontevasti segmentoituva tietojoukko, koska tietokoneissa käytettävissä olevat resoluutiot ovat vakioita. Joskus taas tietojoukolle järkevien segmenttien löytäminen voi olla haastavaa. Näissä tapauksissa on hyvä luoda muutamia rinnakkaisia segmentointeja, joita vertailemalla pystytään etenemään kohti hyvää segmentointimallia. (Croll & Sean, 2009, ss. 57-58)

Segmentoinnissa on mahdollista hyödyntää hahmolakeja. Hahmolait ovat periaatteita, jotka pyrkivät kertomaan millä tavoin ihmisen aivot käsittelevät havaitsemaansa tietoa. Hahmolait perustuvat ihmisen synnynnäisiin ominaisuuksiin hahmottaa ja yhdistellä asioita. Useimmat hahmolaeista on määritelty etenkin näköaistiin perustuen, mutta hahmolakeja voidaan luoda yhtäläisesti esimerkiksi kuuloaistille. Tietotekniikassa ja visualisointien suunnittelussa hahmolakeja voidaan hyödyntää parantamaan luettavuutta ja ymmärrettävyyttä.

Useimmin käytetyt hahmolait ovat luonnostaan ja tiedostamatta käytetyt kokonaisuuden ja tuttuuden lait. Ihmiselle on luonnollista käsitellä pienetkin yksityiskohdat ensin suurempana kokonaisuutena. Kokonaisuuden lakia voidaan hyödyntää visualisoinneissa antamalla käyttäjälle mahdollisimman selkeä ja helposti omaksuttava kokonaiskuva visualisointinäköymän sisällöstä. Tuttuuden lain mukaan ihmiselle hahmottuvat ensin tutut kuviot ja muodot. Monipuolisia ja monimutkaista tietoa käsittelevissä visualisoinneissa tulee pyrkiä käyttämään yksinkertaisia kuvioita, koska niiden hahmottaminen on helpompaa. (Laine, 2004)

4.2.6 Geologisen paikan visualisointi

Käyttäjän paikantaminen ja erilaiset karttoihin perustuvat visualisoinnit ovat viimeisen muutaman vuoden aikana kasvattaneet suosiotaan. Syynä tähän ovat paikannuksen parantunut tarkkuus ja paikannuslaitteiden yleistymisen päätelaitteissa. W3C⁴⁷ on määritellyt virallisen määrittelyn (Geolocation API) päätelaitteen sijain-

⁴⁷<http://www.w3.org/>

nille ja sen käytölle. Geolocation API mahdollistaa asiakasohjelman paikkatiedon jakamisen verkkopalveluissa oleville sovelluksille. (Vrt. Popescu, 2009)

Paikkatiedon keräämisessä tulee ottaa huomioon käyttäjän yksityisyyteen liittyvät tekijät. Mozilla Firefoxin toteutus Geolocation API:sta kysyy joka kerta käyttäjältä luvan paikkatiedon jakamiseen verkkopalvelun kanssa. Verkkopalvelun tehtävänä on taata, että käyttäjä ymmärtää mihin hänen paikkatietoaan käytetään ja tarpeen mukaan hävittää nämä tiedot käytön jälkeen, jotta niitä ei voida käyttää hyväksi.

4.2.7 Haasteet visualisointien toteutuksessa verkkopalveluun

Verkkopalvelujen käytön seurantatiedon visualisointiin ja näkyväksi tekemiseen liittyy useita haasteita. Verkkopalvelut sisältävät niille yksilöllisiä ominaisuuksia tai ominaisuuksien joukkoja, joiden yhdistäminen visualisoinnissa on vaikeaa. Verkkopalveluille yksilölliset ominaisuudet aiheuttavat visualisointien luontiin ongelman, jolloin visualisointeja suunniteltaessa ei voida usein hyödyntää suoraan mitään valmiita toteutuksia. Tämän johdosta verkkopalveluun toteutettavat visualisoinnit täytyy usein suunnitella ja toteuttaa puhtaalta pöydältä.

4.3 Visualisoinnin periaatteet

Tuotettaessa käytön seurantatietoa ihmiselle luettavaan muotoon kohdataan helposti useita haasteita. Miten pystytään esittämään tieto lukijalle ymmärrettävästi. Tulkintaa helpottaa, jos tiedon lukija tietää tulkitessaan esimerkiksi verkkopalvelun kävijämäärää kuvaavaa kuvaajaa:

1. mitä tietoa kuvaaja esittää
2. mistä tieto on peräisin ja
3. mitä tietoa on mahdollisesti jätetty kuvaajan ulkopuolelle.

Kuvaajan toteuttavalla verkkopalvelun tekijälle nämä asiat saattavat vaikuttaa itsestään selviltä, mutta esimerkiksi kävijämääriä kuvaavasta kuvaajasta on lukijan mahdotonta sanoa onko hakukonerobotit mukana tuloksissa. Tällaisissa tilanteissa visualisointina toimivan kuvaajan tulkinta jää epäselväksi.

4.3.1 Luettavuus

Käytön seurantatiedon visualisoinnissa on tarkoituksen mukaista pyrkiä tiedon luettavuuteen. Kuvaajasta tulisi ilmetä mitä tietoa se sisältää ja mitä kuvaajassa olevat arvot tarkoittavat. Tiedon luettavuutta parantavat asteikkoviivat, joiden avulla kuvaajan lukija pystyy helpommin tulkitsemaan kuvaajalla olevia arvoja.

Perustoiminnallisuus, joka on useissa visualisoinneissa tarpeen, on lisätiedon tarjoaminen käyttäjälle osoittamalla tietoa hiirellä. (Kuva 4.4)

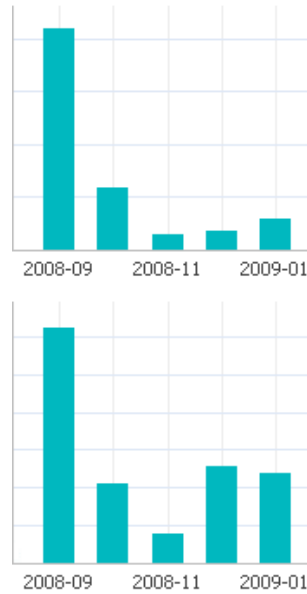


Kuva 4.4: Kuvaajia esitettäessä on hyvä tarjota käyttäjälle mahdollisimman paljon tietoa tiedosta.

Yksittäiseen kuvaajaan ei yleensä ole tarkoituksen mukaista pyrkiä sijoittamaan useata tietosarjaa, koska tämä johtaa usein vaikeasti luettavaan lopputulokseen. Kuitenkin jos tietosarjat ovat vertailukelpoisia ja niiden vertailu on mielekästä, on niiden sijoittaminen samaan kuvaajaan jopa suotavaa. Useita tietosarjoja sisältävissä kuvaajissa tulee huomio kiinnittää tietosarjojen selkeään nimeämiseen ja erottamiseen toisistaan. Nimeämisessä voidaan käyttää värikoodausta.

4.3.2 Asettelu

Asettelulla tarkoitetaan, miten tiedon visualisoinnit asetellaan toisiinsa nähden visualisointinäkyvässä. Asettelu on tärkeää, kun halutaan pystyä esittämään kerralla jokin verkkopalvelun ilmiö tai ilmiöitä. Asettelu voidaan esimerkiksi tehdä erilaisten toimintojen perusteella; eli asettelemalla samaan toimintokokonaisuuteen liittyvät visualisoinnit ja tiedot lähelle toisiaan. Asettelua mietittäessä on hyvä huomata, että allekkain olevissa kuvaajissa pystytään vertailemaan keskenään helpommin x-akselia, kun taas vierekkäin olevissa kuvaajissa y-akselin keskenään vertailu on helpompaa. Kuvassa 4.5 nähdään miten verkkopalvelun kaksi toimintoa on kuvattu erillisissä kuvaajissa. Niiden vertailu samalta aikajaksolta on helppoa, koska x-akselilla kulkeva aika on pystysuuntaisesti samassa kohdassa.



Kuva 4.5: Kun visualisoinnit esitetään allekkain samalla aikajanalla on kuvaajien arvojen vertailu helppoa samoilta x-akselin arvoilla.

Kun useiden tietosarjojen sijoittaminen samaan kuvaajaan on joko teknisesti tai visuaalisesti haasteellista, voidaan tilannetta parantaa visualisointien hyvällä asettelulla. Esimerkiksi jos tietosarjat ovat aikariippuvaisia, voidaan kuvaajat sijoittaa allekkain, jolloin niiden x-akselilla kulkeva aika on kuvaajien kesken linjassa. Näin kuvaajien katsoja pystyy helposti vertailemaan tietosarjojen kuukausittaisia arvoja.

4.3.3 Asteikko

Asteikon valinta on tärkeä osa tiedon esittämistä kuvaajalla. Asteikon valinta riippuu minkälaista tietoa käsitellään sekä tiedon laadusta. Tieto voi olla laadultaan:

- nimellistä (nominal)
- järjestettyä (ordinal)
- arvoalueille asettuvaa (interval) tai
- suhteellista (ratio).

Nimelliselle tiedolle on ominaista, että sitä voidaan käyttää vain tiedon nimeämiseen. Nimellistä tietoa ei voida laittaa matemaattisesti järjestykseen suuruuden tai paremmuuden mukaan. Huomioitavaa on, että numerot voivat olla laadullisesti nimellistä tietoa. Nimellistä tietoa on esimerkiksi luettelo hedelmistä. Järjestetyn tiedon ominaisuus on, että tietoalkioiden välillä voidaan sanoa kumpi tulee ennen toista.

Listan alkioiden järjestysnumero on esimerkki järjestetyn tiedon esiintymästä. Arvoalueen käyttö on ominaista kun kuvataan aikaa. Kuvattaessa päivä arvoalueen alkupiste on päivän alussa ja loppupiste päivän lopussa. Päivä on siis arvoalue tietystä ajasta. Suhteellisella asteikolla voidaan kuvata asioiden keskinäistä suhdetta kuten, että kohde A on kolme kertaa niin suuri kuin kohde B. Suhteellisessa asteikossa ei voida siis sanoa kuinka suuria tietoalkiot ovat, mutta niitä voidaan keskenään vertailla. Esimerkiksi raha on suhteelliselle asteikoille asettuvaa tietoa. (Stevens, 1946, ss. 677-680)

Verkkopalveluissa nämä neljä mittaustasoa voidaan kaventaa kolmeen tasoon, jotka ovat:

- kategorisoitu-
- kokonaisluku- ja
- reaalityttö-

Kategorisoitu tieto vastaa Stevensin kuvaamaa nimellistä tietoa, kokonaisluku- ja reaalityttö- tieto on yhdistelmä Stevensin kuvaamista arvoalueille asettuvasta ja suhteellisesta tiedosta. (Ware, 2004, s. 24-25)

Verkkopalvelujen visualisoinneissa kategorisoituja asteikkoja voidaan käyttää, kun kuvataan tietoa kokonaisuutena. Esimerkiksi kuvattaessa eri käyttäjäröolien lukumäärää verkkopalvelussa on kyse kategorisoidusta tiedosta, jota ei voida laittaa järjestykseen eikä niiden välille voida piirtää yhteyttä.

Usein on mielenkiintoista kuvata tietosisältöjen kehitystä ajan funktiona. Tällöin valitaan aikaväli, jolla tarkastelu tehdään. Aikaväli voi olla päivittäinen, kuukausittainen tai vuosittainen. Aikavälit voidaan asettaa asteikolla järjestykseen.

4.3.4 Luotettavuus

Visualisoinnin luotettavuus on tärkeä tekijä kun tietoa esitellään. Visualisointia lukevan henkilön on pystyttävä luottamaan esitettävän tiedon oikeellisuuteen. Visualisoinnin luotettavuutta voidaan parantaa visualisoinnin läpinäkyvyydellä. Läpinäkyvyyttä voidaan tuoda kertomalla miten visualisointi on luotu ja millä tavoin visualisoinnissa olevaa tietoa on käsitelty ennen visualisointia. Tiedot siitä, mitä tietoa visualisoinnissa on käytetty ja mitä tietoa on jätetty visualisoinnin ulkopuolelle, parantavat visualisoinnin luotettavuutta.

4.3.5 Määrä vs. kertymä

Kuvattaessa tiedon määrää aikajanalla esimerkiksi kuukauden paloissa on kaksi mahdollisuutta esittää tieto. Ensimmäinen voidaan kuvata tiedon kokonaismäärän kehitystä.

Tässä tapauksessa tiedon peräkkäisillä arvoilla on riippuvuussuhde toisiinsa nähden ja luonteva kuvaajatyyppejä on viivadiagrammi. Kertymäkuvaaja antaa kuvan siitä, millä tavoin esitettävä asian määrä on kehittynyt ajan funktiona. Toinen vaihtoehto on esittää määrän kehitys aikaväleittäin. Esitettäessä määrää voidaan nähdä helposti millä aikavälillä tieto on kasvanut eniten ja mikä on esimerkiksi keskimääräinen taso tiedon määrän muutoksessa.

4.4 Visualisointityökalut

Visualisointityökaluilla tarkoitetaan sellaisia ohjelmia tai järjestelmiä, jotka tuottavat niille annetusta tietojoukosta esimerkiksi graafisessa muodossa olevan esityksen. Visualisointityökalun tehtävä on luoda tiedosta esitys, joka on käyttäjän tulkittavissa. Esitysmuoto voi olla esimerkiksi diagrammi tai graafi (katso aliluku 4.2.3). Visualisointityökalu voi kuitenkin tarkoittaa tiedon pukemista sanalliseen esitysmuotoon. Oleellista visualisointityökalulle on, että se tuottaa lisäarvoa olemassa olevan tietojoukon analysointiin.

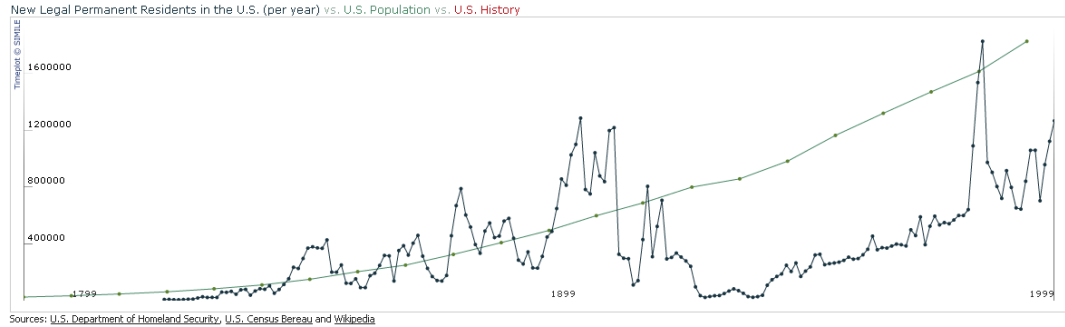
Verkkopalvelujen käytön seurantatiedon visualisointiin soveltuvia työkaluja on olemassa useita. Tärkeää visualisointityökalun valinnassa on määrittellä, missä visualisointityökalun tulee toimia. Halutaanko, että tietoa visualisoidaan reaaliaikaisesti käyttäjän sitä pyytäessä vai tehdäänkö esimerkiksi kuukausittaisia raportteja. Tärkeä valintaperuste työkalulle on, pystytäänkö sille välittämään järjestelmästä saatava käytön seurantatieto, eli voidaanko käytön seurantatiedon tietomalli koostaa sellaiseen muotoon, että visualisointityökalu pystyy käsittelemään sitä.

4.4.1 SIMILE Widgets

SIMILE⁴⁸ (Semantic Interoperability of Metadata and Information in unLike Environments) Widgets on kokoelma avoimeen lähdekoodiin perustuvia vimpaimia (Widgets), jotka on toteutettu Massachusetsin teknillisessä yliopistossa (MIT, Massachusetts Institute of Technology) SIMILE-projektin sivutuotteena. Näihin vimpaimiin kuuluvat tässä tarkemmin esitettävän Timeplotin ohella Timeline, Exhibit ja Runway. Yhteisestä nimityksestä huolimatta vimpaimia voidaan käyttää erikseen ja toisistaan riippumattomasti.

Käytön seurannasta saatavan tiedon näkökulmasta TimePlot on SIMILE Widgets -paketista hyödyllisin, koska se kuvaa havainnollisesti määrällistä tietoa ajan yli esitettynä (kuva 4.6). Tämän kaltainen määrällisesti ajan yli esitettävä tieto on hyvin ominaista suurelle osalle juuri käytön seurannasta saatavaan tietoon.

⁴⁸<http://simile.mit.edu/>



Kuva 4.6: Timeplot on SIMILE Projectin luoma avoimeen lähdekoodiin perustuva visualisointityökalu. (*Timeplot; Web Widget for Plotting Time Series*, s.a.)

Timeplot⁴⁹ hyödyntää DHTML- ja AJAX-teknologiaa eli se pystyy dynaamisesti reagoimaan käyttäjän toimiin, kuten näyttämään halutun tietoalkion tarkan arvon, kun hiiri viedään sen päälle. Timeplottiin voidaan tuoda enemmän Time-linelle ominaisia päivämäärään sidottuja tapahtumia, jolle voidaan antaa kuvaus sekä alkamis- ja päättymispäivä. Näiden avulla Timeplot-kuvaajiin voidaan liittää näppärästi tapahtumia, jolloin kuvaajista voidaan helposti osoittaa oleellisia kohtia.

Tietoalkiot välitetään Timeplotille CSV-muodossa (Comma-Separated Values), jolloin ensimmäisenä arvona annetaan aika käyttäen ISO 8601 -muotoa⁵⁰ (esimerkiksi 2010-03-28). Seuraavat arvot kuvaavat tietoalkioiden arvoja, joiden perusteella Timeplot piirtää kuvaajan. Timeplot pystyy käsittelemään useamman tietoalkiosarjan samassa kuvaajassa sekä esittämään lisäksi näiden rinnalla erikseen tuotavassa tiedostossa tietosarjoihin liittyviä tapahtumia (kuva 4.6). CSV-tiedosto välitetään Timeplotille käyttäen `Timeplot.loadText()`-funktion avulla, jolle annetaan parametrina CSV-tiedoston sijainti. CSV-tiedosto on mahdollista luoda dynaamisesti pyynnön yhteydessä tai se voidaan luoda ajoitetusti esimerkiksi kerran tunnissa.

4.4.2 Ext JS

Ext JS⁵¹ on selainriippumaton JavaScript-kirjasto, jota hyödynnetään laajasti yrityksissä⁵². Ext JS mahdollistaa jopa kokonaisten verkkopalvelujen toteuttamisen mutta usein sitä hyödynnetään vain soveltuvin osin. Ext JS on monipuolinen kirjasto, jonka avulla voidaan toteuttaa dynaamisia ja rikkaita verkkopalveluita. Perustoiminnallisuuksiin kuuluvat välilehdet, taulukot, graafit ja ikkunointi. Kirjastosta löytyy työkalut lomakkeiden, työkalupalkkien ja valikoiden luomiseen. Ext JS:stä on saatavilla sekä kaupallinen että avoimeen lähdekoodiin perustuva ratkaisu.

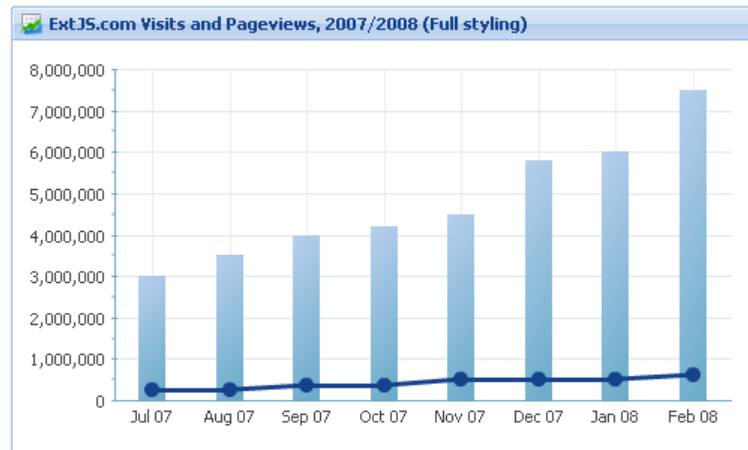
⁴⁹<http://www.simile-widgets.org/timeplot/>

⁵⁰ISO 8601 Data elements and interchange formats – Information interchange – Representation of dates and times. (1st ed.) International Organization for Standardization, 15.6.1988

⁵¹<http://www.extjs.com/>

⁵²<http://www.extjs.com/company/customers.php>

Käytön seurannasta saatavan tiedon kannalta Ext JS:n kiinnostavimmat osat löytyvät taulukko- ja graafityökaluista. Ext JS:lle voidaan helposti välittää sekä staattista että dynaamista tietoa ohjelmointikielen muuttujien avulla. Tietoalkioille voidaan antaa sarakekohtaiset nimet, joten niiden hallinnointi ja visualisointi on havainnollista, kun voidaan toimia indeksinumeroiden sijaan oikeilla nimillä. (Kuva 4.7)



Kuva 4.7: Ext JS on JavaScript kirjasto, jolla voidaan tehdä hyvin monipuolisia ja rikkaita verkkopalveluja (*Ext JS; Charts, s.a.*)

4.4.3 Taulukkolaskentaohjelmistot

Taulukkolaskentaohjelma on graafisessa käyttöjärjestelmässä toimiva ohjelma, joka käsittelee tietoa taulukkomuodossa. Tiedot asetetaan taulukossa soluihin ja ne voidaan jakaa soluissa loogiseen järjestykseen. Pystysuuntaista solujen joukkoa nimitetään sarakkeeksi ja vaakasuuntaista riviksi. Suosituimmat taulukkolaskentaohjelmistot ovat Microsoft Excel ja OpenOffice.org Calc.

Kun tietojoukko on tuotu taulukkolaskentaohjelmaan, voidaan sitä visualisoida järjestelmän avulla. Jos käsiteltävänä on esimerkiksi tietojoukko, jossa on avain-arvo -pareja, joissa avain kertoo päivämäärän ja arvo päiväkohtaisen kävijämäärän verkkopalvelussa voidaan avain-arvo -parit asettaa taulukkolaskentaohjelmaan niin, että sarakkeeseen "A" asetetaan allekkain päivämäärät ja sarakkeeseen "B" päivämääriä vastaavat kävijämäärät. Sarakkeet on taulukkolaskentaohjelmissa nimetty aakkosten avulla, että ensimmäinen sarake on nimeltään "A". Kuvatusta tietojoukosta on mahdollista piirtää viivadiagrammi, jossa x-akselilla kuvataan juokseva päivämääränumerointi ja y-akselilla päiväkohtaiset kävijämäärät. Taulukkolaskentaohjelmaan tuotua tietojoukkoa voidaan helposti visualisoida usein eri tavoin ja kokeilla minkä tyyppinen visualisointi olisi paras. Edellä kuvatussa esimerkissä voitaisiin vertailla palkki- ja viivadiagrammin paremmuutta tiedon esittämisen välineenä.

Joustavuus ja diagrammien monipuolisuus ovat taulukkolaskentaohjelmien selkeitä etuja suhteessa moniin muihin järjestelmiin. Taulukkolaskentaohjelmien heikkoutena on, että ne eivät ole dynaamisia eivätkä integroidu suoraan seurattavana olevaan verkkopalveluun. Haasteena on saada seurattavasta järjestelmästä vietyä sellainen tietomalli, joka voidaan tuoda järjestelmään sen ymmärtämässä muodossa. Tiedon vientiä verkkopalvelusta käsitellään laajemmin aliluvussa 3.8.1. Taulukkolaskentaohjelmien monimutkaisuus voi vaikeuttaa niiden käyttämistä käytön seurattatiedon visualisoinnissa.

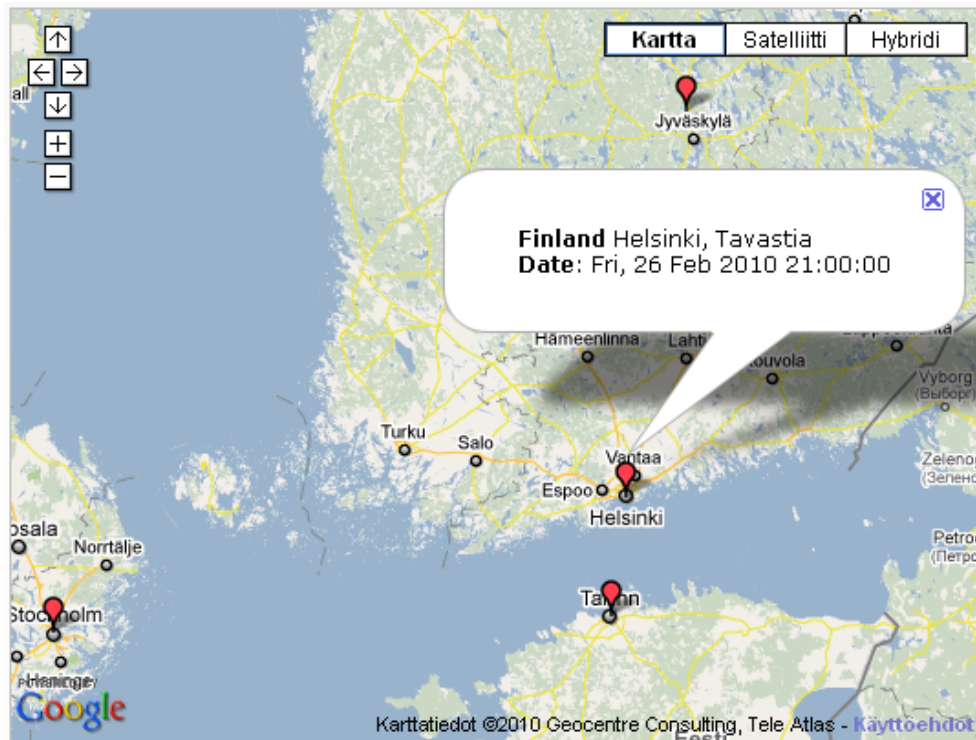
4.4.4 Google Maps

Sijainteihin liittyvien visualisointien takana on vahvana tekijänä Googlen toteuttama Google Maps -karttapalvelu⁵³. Google Maps on verkkopalvelu, joka tarjoaa käyttäjille kattavat karttatiedot. Google Maps -palveluun on mahdollista tuottaa omia karttanäkymiä populoimalla karttoja erilaisilla sijaintitiedoilla (kuva 4.8). Google Maps palvelun avulla on siis mahdollista luoda verkkopalvelun käyttäjistä erilaisia visualisointeja perustuen käyttäjien paikkatietoon. Paikkatieto voidaan perustaa joko käyttäjän verkkopalveluun asettamaan sijaintiin. Vaihtoehtoisesti käyttäjän asiakasohjelma voidaan paikantaa IP-osoitteen avulla. Mobiililaitteissa paikannus voidaan tehdä satelliittisignaalien tai tukiasematietojen perusteella.

Kuvassa 4.8 näytetään miten yksittäisen yhtyeen keikkakalenteri on mahdollista visualisoida Google Maps API:n⁵⁴ (Application Programming Interface) avulla. Mahdollisesti vaikeasti luettava lista keikoista aikajärjestyksessä pystytään esittämään paikkaan sidottuna karttana, joka tukee käyttäjän havainnointia.

⁵³<http://maps.google.com/>

⁵⁴<http://code.google.com/apis/maps/>



Kuva 4.8: Google Mapsilla voidaan luoda sijaintitietojen perusteella esimerkiksi visualisointi halutun artistin tulevista keikoista.

Google Maps tarjoaa kehittäjille monipuoliset työkalut visualisointien tekemiseen ja karttatiedot ovat siinä määrin kattavat, että sijaintitieto voidaan palvelussa esittää hyvinkin tarkasti. Google Mapsin alati kehittyvällä katunäkymäominaisuudella voidaan näyttää sijaintitieto todellisessa ympäristössään. Uusin versio rajapinnasta on kirjoitushetkellä (20.2.2010) järjestyksessään 3. Versio uudistaa palvelua, että karttavisualisoinnit voidaan yhä helpommin sisällyttää työpöytäympäristöihin ja mobiililaitteisiin. Edelliset versiot ovat pääasiassa toimineet WWW-ympäristöissä ja vaatineet toimiakseen modernin selainohjelmiston.

4.4.5 Visualisointiympäristön valinta

Visualisointityökalut ovat väline, jolla kerätty käytön seurantatieto voidaan tehdä ihmisnäölle havainnolliseksi. Tästä johtuen työkalulla, joka visualisoinnin tuottaa on ensiarvoisen tärkeä rooli prosessissa. Työkalun valinnassa on tärkeää, että sen avulla pystytään tuottamaan sellaisia visualisointeja, jotka ovat ominaisia kerättävänä olevalla tietojoukolla. Työkalun tulee olla sellainen, että siihen voidaan ylipäänsä syöttää käsittelyssä oleva tietojoukko. Hyviä käytön seurantatiedon visualisointityökalun ominaisuuksia ovat nopea tiedon prosessointi, helppokäyttöisyys, korkea mukautuvuus erilaisiin tietojoukkoihin ja visualisointien selkeys. (Vrt. Kaushik, 2010, ss. 17-19)

4.5 Visualisoinnin jakelu

Visualisoinnin saavutettavuus on tärkeä tekijä, koska kerätty käytön seurantatieto tavoittaa ne ihmiset, jotka tietoa käyttävät. Saavutettavuus liittyy osaltaan oikean visualisointityökalun valintaan, koska saavuttaakseen tiedon on havaittajan osattava käyttää välinettä, jolla tieto visualisoidaan. Toisaalta saavutettavuuteen liittyy käytön seurantatiedon vienti verkkopalvelusta, jota käsitellään tarkemmin aliluvussa 3.8.1.

4.5.1 Visualisoinnin toteutus verkkopalvelussa

Verkkopalvelun yhteyteen toteutettu visualisointi on usein paras vaihtoehto jos tämä on vain teknisesti mahdollista. Verkkopalvelussa oleva visualisointi toimii tietojoukolle luonnollisessa ympäristössä ja on sulautettavissa sivuston muuhun ulkoasuun sopivaksi. Kun visualisoinnit toteutetaan verkkopalveluun suoraan, on niiden hyödyntäminen verkkopalvelun käyttäjien suuntaan helpompaa ympäristön ollessa sama. Verkkopalveluun toteutettavien visualisointien haasteita ovat olemassa olevien työkalujen puutteelliset ominaisuudet, sekä mahdolliset verkkopalvelun toteutusteknologian asettamat rajoitteet.

4.5.2 Visualisointi verkkopalvelun ulkopuolella

Käytön seurantatieto on mahdollista viedä verkkopalvelusta verkkopalvelun ulkopuolisiin ohjelmistoihin. Tämän vaihtoehdon vahvuutena on yleensä erilaisten ohjelmistojen ja ominaisuuksien kirjon laajuus. Käytön seurantatietoa voidaan prosessoida erilaisissa taulukkolaskentaohjelmissa ja erityisesti suurten tietomäärien prosessointiin erikoistuneissa ohjelmistoissa (katso aliluku 4.4.3). Visualisointien viemisellä verkkopalvelun ulkopuolelle heikkoutena on niiden saavutettavuus. Ulkopuoliset visualisointiohjelmat toimivat paikallisesti, eivätkä ole suoraan saatavilla Internetistä.

5 Käytön havainnointi: Erimenu.fi

Älykkään ja tietoa monipuolisesti keräävän käytön seurantajärjestelmän esittelyn tukena käytetään pääasiassa Erimenu.fi-verkkopalveluun kesän ja syksyn 2009 aikana toteutettua käytön seurantajärjestelmää. Yksittäistapauksissa viitataan muissa projekteissa toteutettuihin ominaisuuksiin. Luvussa esitellään ensin lyhyesti Erimenu.fi-verkkopalvelun keskeisimmät osat, jonka jälkeen kerrotaan miten seurantajärjestelmä verkkopalveluun toteutettiin. Kerätyn tiedon perusteella luotuja visualisointeja esitellään etenkin niiden avulla havaittujen muutosten kautta. Erimenu.fi-verkkopalveluun toteutettu käytön seurantajärjestelmä nimettiin Loggeriksi (Lokittaja).

Tähdennetään, että työn tarkoituksena ei ole selittää verkkopalvelun käytössä tapahtuvaa muutosta, vaan tarkoituksena on pystyä havaitsemaan mahdollinen muutos Työssä ei pyritä kertomaan, että

- verkkopalvelussa käytetään nykyään avainsanoja, koska lähipiiritoiminnallisuus toi ympäristöön yhteisöllisyyden tunnetta.

Tutkimuksen pyrkimys on mahdollistaa seurantatietoa lukevalle henkilölle muutoksen havainnointi, jotta voidaan sanoa, että

- verkkopalvelussa lisätään enemmän avainsanoja sen jälkeen, kun lähipiiritoiminnallisuus otettiin käyttöön verrattuna vastaavaan aikaisempaan ajanjaksoon.

Tämä rajausta johtuu siitä, että halutaan keskittyä juuri muutoksen todentamiseen liittyvään tutkimukseen. Lisäksi muutoksen selittäminen vaatisi laajamittaisempaa käyttäjätutkimusta, joka rajaa tämän tutkimusnäkökulman työn ulkopuolelle.

5.1 Erimenu.fi

Erimenu.fi on Tampereen Teknillisen Yliopiston (TTY) hypermedialaboratorion sekä Pirkanmaan Allergia- ja Astmayhdistys ry:n kehittämä verkkopalvelu. Erimenu.fi-verkkopalvelu palvelee erityisruokavaliota noudattavia henkilöitä, tarjoamalla heille ajantasaista tietoa juuri heille soveltuvista elintarvikkeista. Erimenu.fi:ssä ovat esillä lisäksi elintarvikkeiden tuottajat ja elintarvikkeisiin liittyvät ruokaohjeet. Erimenu.fi on toteutettu Aatu-hankkeen puitteissa pääasiallisesti Raha-automaattiyhdistyksen rahoituksella. Erimenu.fi on toteutettu tulkattavalla Python-ohjelmointikielellä hyödyntäen avointa Django-kehitysympäristöä⁵⁵. Tiettyjä dynaamisia osia on toteutettu JavaScriptillä ja AJAX:lla.

⁵⁵<http://www.djangoproject.com/>

Tietosisältöjen osalta Erimenu.fi:n oleelliset osat ovat verkkopalveluun lisätyt ainesosat, tuotteet, tuottajat ja ruokaohjeet. Ainesosat ovat peruselintarvikkeita, joille on määritelty aineeseen liittyvät perustiedot. Tuotteet ovat ainesosista koostettuja ainesosakokonaisuuksia, joita on mahdollista hankkia. Tuottajat ovat tuotteiden valmistajia. Ruokaohjeet ovat tuotteista ja ainesosista koostettuja kokonaisuuksia, jotka määrittelevät miten jokin ruoka valmistetaan.

Käyttäjälle Erimenu.fi-verkkopalvelun olennaisimmat ominaisuudet ovat käyttäjän oma lista ja ruokavalio. Käyttäjän oma lista on toiminnallisuus, joka mahdollistaa käyttäjälle verkkopalvelussa olevien tuotteiden ja ruokaohjeiden lisäämisen käyttäjän henkilökohtaiselle listalle. Lista toimii periaatteessa verkkopalvelussa kirjainmerkkien (bookmarks) tapaan. Ruokavalio toiminnallisuudessa käyttäjän on mahdollista määritellä verkkopalveluun oma henkilökohtainen ruokavalionsa. Määritellyn ruokavalioon perustuen käyttäjä saa tietoa hänelle sopivista tuotteista ja ruokaohjeista. Lisäksi määritelty ruokavalio voidaan jakaa muiden verkkopalvelun käyttäjien kesken.

Lähipiiri on Erimenu.fi-verkkopalveluun toteutettava toiminnallisuus, jonka avulla käyttäjien on mahdollista merkitä toisia Erimenu.fi-käyttäjiä omiin Lähipiireihinsä. Lähipiirin toteuttaa useista sosiaalisista verkkopalveluista, kuten suositusta Facebook-palvelusta⁵⁶, löytyvän kaveriksi merkkaisominaisuuden. Toiminnallisuuden nimeksi valittiin Lähipiiri, koska nähtiin, että verkkopalvelussa Lähipiiriin merkittävät kaverit eivät välttämättä suoraan ole käyttäjän kavereita tai ystäviä reaali maailmasta vaan heidän ruokavalionsa ovat lähellä toisiaan. Näin käyttäjät voivat olla toisensa Lähipiireissä kuitenkin assosioitumasta suoraan ystäviksi. Nähtiin siis, että Lähipiiri olisi terminä neutraali ja tämän oletettiin alentavan kynnystä ominaisuuden käyttöönottamiselle käyttäjien keskuudessa. Lähipiiriin lisäämisen lisäksi käyttäjän on mahdollista jakaa Lähipiirissään olevat käyttäjät haluamiinsa ryhmiin. Käyttäjien jakaminen ryhmiin on Facebookissa käytetty ominaisuus. Ryhmä on Lähipiirissä olevien käyttäjien osajoukko. Lähipiiriin lisättyjä käyttäjiä ja heille määriteltyjä ryhmiä nähdään kuvassa 5.1.

Erimenu.fi:ssä käyttäjien välille luotava Lähipiiri on kaksisuuntainen. Kaksisuuntaisuudella tarkoitetaan, että molempien käyttäjien on hyväksyttävä Lähipiirisuhde, jotta Lähipiiri-yhteys käyttäjien välille syntyy. Kaksisuuntaisesti luotavaa yhteyttä suositaan luotaessa yhteys kahden niin sanotun todellisen tai tietoisien objektien välille, jotka tässä tapauksessa ovat verkkopalvelun käyttäjiä. Yksisuuntaisia yhteyksiä Erimenu.fi-verkkopalvelussa käytetään, kun käyttäjä lisää verkkopalvelussa olevia tuotteita omalle listalleen. Kun käyttäjä lisää tuotteen omalle listalleen, tuotteen ei tarvitse hyväksyä listalle lisäämistapahtumaa.

Lähipiirin avulla Erimenu.fi-verkkopalvelun käyttäjää pyritään ensisijaisesti

⁵⁶<http://www.facebook.com/>

tukemaan mielenkiintoisten käyttäjien löytämisessä verkkopalvelusta. Lähipiirissä olevat käyttäjät listataan verkkopalvelussa selkeästi omalla sivullaan, josta käyttäjä pystyy helposti siirtymään haluamansa käyttäjän sivulle. Lähipiiriin perustuen käyttäjän on mahdollista määritellä omaa näkyvyyttään verkkopalvelussa. Käyttäjän on mahdollista rajata omien tietojensa näkyvyys vain omalle Lähipiirillensä. Käyttäjän ruokavalion näkyvyyden määrittely on vielä mahdollista ryhmäkohtaisesti. (Kuva 5.1)

Käyttäjä	Ryhmät
 » Poista	Hlab » Muokkaa
 » Poista	Turku, ex-Hlab » Muokkaa
 » Poista	Koti » Muokkaa
 » Poista	Hlab » Muokkaa
 » Poista	Hlab » Muokkaa
 » Poista	Hlab » Muokkaa

Kuva 5.1: Lähipiiriin on mahdollista lisätä muita Erimenu.fi-verkkopalvelun käyttäjiä. Käyttäjien yksityisyyteen liittyvät tiedot on sumennettu.

5.2 Sivulatausten seuranta Erimenu.fi:ssä

Erimenu.fi-verkkopalvelussa käytön seurantajärjestelmä on toteutettu “middleware”-tasolle. Middleware-taso on Djangoissa sivulatauspyyntöihin ja vasteisiin kytkeytyvä taso, jolla pystytään kevyesti vaikuttamaan sivuston sisäänmenoihin ja ulostuloihin. Middleware-tasolla käytön seurantajärjestelmän käytettävissä ovat kaikki sivulatauspyyntöön liittyvät muuttujat. (*Middleware*, 2010)

Sivulatauksiin liittyvien tietojen tallennus toteutettiin Erimenu.fi-verkkopalvelussa perustuen näkymien URI-tunnisteisiin. Tähän ratkaisuun päädyttiin, koska Django-kehitysympäristö mukaillee malli-näkymä-ohjain -arkkitehtuuria (katso aliluku 2.3). Näin sivulataukset ovat yksilöitävissä tiettyihin osoiteavaruuksiin ja ohjaintasolla pystytään määrittelemään mitä tietoja on tallennettavissa.

Erimenu.fi-verkkopalveluun toteutettu käytön seurantaohjelmisto tunnistaa sivulatauspyynnön osoitteesta, mitä tietoja sivulatauspyynnöstä voidaan tallentaa ja kerää tämän perusteella sivulataukseen liittyvät tiedot. Esimerkiksi tuotesivulla tunnistetaan sivulatauspyynnöstä tuotteen nimi ja sen perusteella haetaan tuotteen

yksilöivä ID-tunnistearvo (Identification (ID) value) tietokannasta. Tämä tieto yhdistettynä muihin sivulatauksesta kerättäviin tietoihin tallennetaan tietokantaan uutena rivinä.

Sivulatauksen tiedot tallennetaan tietokantaan siten, että jokaiselle tiedolle on varattu oma sarake. Vaihtoehtoisesti sivulatauksesta mahdollisesti tallennettavat tiedot voitaisiin kerätä nimettyinä arvoina yhteen tietokantasarakkeeseen seuraavasti: "arvon_nimi1: arvo1<erotin>arvon_nimi2: arvo2". Tällä tavalla voitaisiin minimoida käsiteltävien sarakkeiden määrä ja rakentaa tietokannan rakenne, jos käsiteltävä järjestelmä olisi laajempi ja sisältäisi enemmän vain harvoin tallennettavia arvoja. Tietojen tallentaminen yhteen sarakkeeseen lisää tiedon jalostamisen tarvetta käytön seurantatietoa visualisoitaessa, joten Erimenu.fi:n laajuisessa järjestelmässä on vielä perusteltua antaa jokaiselle tallennettavalla arvolle oma sarakkeensa tietokannassa.

Sivulatausten seurantaan Erimenu.fi-verkkopalvelussa oli toteutettu ennen tätä projektia toiminnallisuus, joka seurasi verkkopalvelussa olevia tuottaja- ja tuotesivuja. Tätä toiminnallisuutta hyödynnettiin, kun haluttiin kertoa elintarvikkeiden tuottajille heidän tuotteidensa kävijämääriä.

Erimenu.fi-verkkopalvelussa määriteltiin sivulatauksista tallennettaviksi tietokantarakkeiksi:

Taulukko 5.1: Sivulatauksen yhteydessä tallennettavat reseptiin liittyvät tiedot.

Tallennettava tieto	Polku	Selitys
Resepti ID	/ruokaohje/{% recipe %}/	Tallentamalla sivulatauksen yhteydessä reseptin yksilöivä ID-arvo voidaan sivulatausten jatkokäsittelyssä yksilöidä koskemaan esimerkiksi vain haluttuja reseptejä.

Taulukko 5.2: Sivulatauksen yhteydessä tallennettavat tuottajaan liittyvät tiedot.

Tallennettava tieto	Polku	Selitys
Tuottaja ID	/tuote/{% producer %}/	Tallentamalla sivulatauksen yhteydessä tuottajan yksilöivä ID-arvo voidaan sivulatausten jatkokäsittelyssä yksilöidä koskemaan esimerkiksi vain tiettyä tuottajaa, jolloin voidaan tuottaa arvokasta informaatiota tuottajakohteisesti.

Taulukko 5.3: Sivulatauksen yhteydessä tallennettavat tuotteeseen liittyvät tiedot.

Tallennettava tieto	Polku	Selitys
Tuote ID	/tuote/{% producer %}/{% product %}/	Tallentamalla sivulatauksen yhteydessä tuotteen yksilöivä ID-arvo voidaan sivulatausten jatkokäsittelyssä yksilöidä koskemaan esimerkiksi vain tiettyä tuotetta, jolloin voidaan tuottaa arvokasta informaatiota tuotekohtaisesti.

Taulukko 5.4: Sivulatauksen yhteydessä tallennettavat avainsanaan liittyvät tiedot.

Tallennettava tieto	Polku	Selitys
Avainsana ID	/haku/{% tag %}/	Avainsanasivulatauksia seuraamalla pystytään esimerkiksi tutkimaan sitä minkä tyyppisissä tilanteissa avainsanoja käytetään kun yhdistetään avainsanoihin tieto siitä mistä avainsanasivulle on tultu.

Taulukko 5.5: Sivulatauksen yhteydessä tallennettavat lähipiiriin liittyvät tiedot.

Tallennettava tieto	Polku	Selitys
Lähipiiriryhmän ID	/lahipiiri/{% group %}/	Ryhmien ID-arvot tallentamalla voidaan seurata esimerkiksi sitä miten aktiivisesti Erimenu.fi käyttäjät käyttävät hyväkseen lähipiirissä olevat henkilöt ryhmitteleviä ryhmäsivuja.

Taulukko 5.6: Sivulatauksen yhteydessä tallennettavat ruokavalioon liittyvät tiedot.

Tallennettava tieto	Polku	Selitys
Ruokavalio ID	/ruokavalio/{% diet %}/	Ruokavalioiden ID-arvot tallentamalla voidaan seurata sitä miten paljon käyttäjät käyttävät ruokavalioita hyödykseen.

Taulukko 5.7: Sivulatauksen yhteydessä tallennettavat käyttäjäprofiiliin liittyvät tiedot.

Tallennettava tieto	Polku	Selitys
Profiili ID	/kayttaja/{% username %}/ ja /yllapito/kayttajat/{% username %}/	Tallentamalla sivulatauksen yhteydessä profiilin yksilöivä ID-arvo voidaan sivulatausten jatkokäsittelyssä yksilöidä yksittäiset profiilit ja esimerkiksi kuvata niiden suosiota.

Taulukoissa 5.1 – 5.7 ID:llä viitataan kyseisen tietoalkion yksikäsitteiseen numeeriseen tunnistearvoon, joka yksilöi kyseisen tietosisällön tietokannassa, johon tiedot tallennetaan. Polku viittaa URL-tunnisteeseen, jossa kyseistä tietosisältöä käsitellään. Polkua käytetään hyödyksi, kun sivulatauksesta tunnistetaan tallennettavissa olevia tietoja. Lisäksi yleisesti koko järjestelmän tasolla kaikissa näkymissä tallennetaan taulukossa 5.8 esitettävät arvot.

Taulukko 5.8: Sivulatauksesta tallennettavat yleiset ominaisuudet.

Sarakkeen nimi	Arvo
search_key	Hakusana
ip_address	IP-osoite
creator_id	Kirjautuneen käyttäjän ID
post_data	[0 1] - riippuen onko näkymässä asetettu POST-ympäristömuuttuja
files_data	[0 1] - riippuen onko näkymässä lähetty tiedosto.
referer	Viittaavan näkymän URI-osoite
user_agent	Käyttäjän selaimen ja käyttöympäristön tiedot
user_lang	Käyttäjän selaimen kieli ja merkistöstandardi
absolute_uri	Näkymän absoluuttinen osoite
duration	Näkymän ladonnan kesto
exception	Djangon palauttamattomat poikkeukset
created	Sivulataustapahtuman aikaleima

Osa taulujen 5.1 – 5.8 tiedoista ovat sellaisia, että ne tallennetaan jokaisen sivulatauksen yhteydessä (kuten sivulatauksen aikaleima). Osa tiedoista liittyy vain tiettyihin sivulatauksiin (kuten kirjautuneen käyttäjän ID-arvo tallennetaan vain kun käyttäjä on kirjautunut). Tämä tarkoittaa, että tietyt tietokannan sarakkeet asettuvat sivulataustiedon kohdalla arvoon NULL. NULL-arvon käyttö on kannattavaa, koska SQL:ssä voidaan tällöin hyödyntää NULL-arvoon liittyviä komentoja ja siten parantaa kyselyjen tehokkuutta (Vrt. *SQL NULL Values*, 2010).

Kuvassa 5.2 nähdään miten taulukoissa 5.1 – 5.8 esiteltävät ominaisuudet näkyvät MySQL-tietokannanhallintajärjestelmän PhpMyAdmin-hallintaohjelmistossa.

	Sarake	Tyyppi	Aakkosjärjestys	Tyhjä	Oletusarvo	Lisätiedot
<input type="checkbox"/>	id	int(11)		Ei		auto_increment
<input type="checkbox"/>	recipe_id	int(11)		Kyllä	NULL	
<input type="checkbox"/>	producer_id	int(11)		Kyllä	NULL	
<input type="checkbox"/>	product_id	int(11)		Kyllä	NULL	
<input type="checkbox"/>	profile_id	int(11)		Kyllä	NULL	
<input type="checkbox"/>	diet_profile_id	int(11)		Kyllä	NULL	
<input type="checkbox"/>	tag_id	int(11)		Kyllä	NULL	
<input type="checkbox"/>	search_key	varchar(255)	utf8_swedish_ci	Kyllä	NULL	
<input type="checkbox"/>	action	varchar(255)	utf8_swedish_ci	Kyllä	NULL	
<input type="checkbox"/>	ip_address	varchar(255)	utf8_swedish_ci	Kyllä	NULL	
<input type="checkbox"/>	user_id	varchar(255)	utf8_swedish_ci	Kyllä	NULL	
<input type="checkbox"/>	creator_id	int(11)		Kyllä	NULL	
<input type="checkbox"/>	post_data	int(11)		Ei		
<input type="checkbox"/>	files_data	int(11)		Ei		
<input type="checkbox"/>	referer	varchar(255)	utf8_swedish_ci	Ei		
<input type="checkbox"/>	user_agent	varchar(255)	utf8_swedish_ci	Ei		
<input type="checkbox"/>	user_lang	varchar(255)	utf8_swedish_ci	Ei		
<input type="checkbox"/>	absolute_uri	varchar(255)	utf8_swedish_ci	Ei		
<input type="checkbox"/>	duration	decimal(10,2)		Ei		
<input type="checkbox"/>	query_count	int(11)		Ei		
<input type="checkbox"/>	exception	longtext	utf8_swedish_ci	Kyllä	NULL	
<input type="checkbox"/>	created	datetime		Ei		

Kuva 5.2: Sivulataustiedon tallentavan tietokantataulun rakenne.

Kuvassa 5.3 nähdään miten verkkopalvelujen käytön seuranta näkyy tietosisältöjen tallentamisessa tietokantaan.

	Sarake	Tyyppi	Aakkosjärjestys	Tyhjä	Oletusarvo	Lisätiedot
<input type="checkbox"/>	id	int(11)		Ei		auto_increment
<input type="checkbox"/>	title	varchar(255)	utf8_swedish_ci	Ei		
<input type="checkbox"/>	description	longtext	utf8_swedish_ci	Ei		
<input type="checkbox"/>	type_id	int(11)		Ei		
<input type="checkbox"/>	main_substance_id	int(11)		Ei		
<input type="checkbox"/>	default_image_id	int(11)		Kyllä	NULL	
<input type="checkbox"/>	created	datetime		Ei		
<input type="checkbox"/>	creator_id	int(11)		Ei		
<input type="checkbox"/>	last_modified	datetime		Ei		
<input type="checkbox"/>	modifier_id	int(11)		Ei		

Kuva 5.3: Ainesosa-tietosisällön tallentavan tietokantataulun rakenne.

Määriteltäessä sivulatauksesta kerättäviä tietoja tulee ottaa huomioon, että tallennettavat tiedot eivät vaaranna järjestelmän tietoturvaa millään tavoin. Esimerkiksi POST-muuttujassa kuljetetaan käyttäjän kirjautumistietoja, joten muut-

tujan tallentaminen sellaisenaan sivulatauksen yhteyteen tarkoittaisi, että käyttäjän tunnus ja salasana olisivat selkokielisenä luettavissa käytön seurantatiedoista.

Erimenu.fi:ssä POST-ympäristömuuttujan tutkiminen on tärkeää, koska siitä voidaan päätellä, onko käyttäjä suorittanut järjestelmässä jonkun toiminnon eli lähettänyt lomakkeen. Kun sivulatausten perusteella halutaan tutkia miten monta kertaa käyttäjät keskimäärin päivittävät profiilejaan, täytyy seurata miten usein profiilin muokkaussivulla on lähetetty lomake. Profiilin muokkaussivun lataaminen ei yksinään kerro, että profiilia olisi päivitetty. Erimenu.fi-verkkopalveluun ja toteutuksessa hyödynnettyyn Django-kehitysympäristöön⁵⁷ liittyvä yksityiskohta tässä tarkastelussa on, että Django ei välitä POST-ympäristömuuttujassa lomakkeen lähettävän painikkeen (submit button) arvoa (value). Eli jos lomake sisältää POST-ympäristömuuttujan välittämän tietosisällön osalta vain lomakkeen lähettävän painikkeen, näyttää käytön seurantajärjestelmän näkökulmasta, ettei lomaketta ole lähetetty. Erimenu.fi:ssä tämänkaltainen tilanne on esimerkiksi käyttäjäkuvan päivityslomakkeessa, joka sisältää tiedoston valintakentän ja lähetä-painikkeen. Ongelma ratkeaa tarkkailemalla käytön seurantatiedossa FILES-ympäristömuuttujaa, joka välittää käyttäjän lomakkeissa lähettämät tiedostot.

Erimenu.fi:ssä tietojen tallennus tapahtuu kooditasolle, että ensin tarkistetaan sivuston osoitteesta missä palvelun osassa käyttäjä on. Sivuston osan perusteella päätellään mitä tietoja tältä sivulta voidaan tallentaa. Kun käyttäjä on palvelussa osoiteavaruudessa “/haku/*” voidaan päätellä, että sivulatauksesta pystytään normaalien muuttujien lisäksi mahdollisesti tallentamaan hakusana tai avainsanan ID-arvo.

5.3 Erimenu.fi-verkkopalvelusta tuotettu lokitieto

Erimenu.fi:ssä käytön seurantatiedon vienti toteutettiin kertomalla järjestelmälle mistä verkkopalvelun osasta (eli toiminnallisuudesta) tiedon haluaa viedä. Vientiin sisällytettiin mahdollisuus valita miltä aikaväliltä tietojoukon haluaa olevan kerätty. Nämä ominaisuudet helpottavat tiedon jälkikäsitteilyä, kun voidaan rajata tieto koskemaan vain haluttua osaa verkkopalvelusta halutulla ajalla. Tämä ratkaisee osaltaan sitä ongelmaa, mikä saattaa muodostua lokitietomäärien massiivisesta lukumäärästä. Lokitiedon sisältävä tiedosto palautetaan selaimessa käyttäjälle ladattavana tiedostona, jolloin käyttäjä voi päättää mihin ja millä nimellä hän haluaa tiedon tallentaa. Mahdollista olisi tulostaa lokitieto suoraan verkkopalvelun näkymään, josta se olisi kopioitavissa, mutta lokitiedon määrän suuruudesta johtuen tulostaminen voi olla selainohjelmistolle erittäin raskasta.

Kuvassa 5.4 nähdään Erimenu.fi-verkkopalveluun toteutettu lokitiedon vienti-

⁵⁷<http://www.djangoproject.com/>

lomake, josta on mahdollista määrittellä mistä verkkopalvelun tietosisällön osasta vientitiedon haluaa luoda. Lokitieto voidaan viedä järjestelmästä rivitettynä ja sisennettynä, jolloin se on ihmissilmälle helpommin luettavissa tai yhdelle riville tulostettuna, jolloin sen käsittely on nopeampaa.

Kuva 5.4: Lokitiedon viennin toteuttava lomake Erimenu.fi:ssä.

Alla esimerkki järjestelmästä viedystä lokitiedosta, joka sisältää kaikki yksittäisestä sivulataustapahtumasta tallennetut tiedot. Tallennettujen tietojen kuvaukset löytyvät taulukoista 5.1 – 5.8.

```
<?xml version=1.0 encoding=utf-8?>
<djangoexport>
<ActivityLogEvents>
<id> 1 </id>
<ip_address> 130.230.8.13 </ip_address>
<user_id> MTMwLjZzMC44LjEz </user_id>
<creator> 5 </creator>
<post_data> 0 </post_data>
<get_data> 0 </get_data>
<referer> http://tebest.dev.ee.tut.fi/erimenu/kirjaudu/ </referer>
<user_agent>
  Mozilla/5.0 (Windows; U; Windows NT 5.1;
  en-US; rv:1.9.1.3) Gecko/20090824 Firefox/3.5.3
  (.NET CLR 3.5.30729)
```

```

</user_agent>
<user_lang> en_US.UTF-8 </user_lang>
<absolute_uri> http://tebest.dev.ee.tut.fi/erimenu/ </absolute_uri>
<duration> 5.12 </duration>
<query_count> 598 </query_count>
<created> 2009-10-02 11:14:45 </created>
</ActivityLogEvents>
</djangoexport>

```

5.4 Muutoksen havainnointi

Erimenu.fi-verkkopalveluun toteutettiin noin 30 erilaista verkkopalvelun käyttöä seuraavaa visualisointia. Visualisoinnit toteutettiin erilaisina kuvaajina ja mittareina. Näiden joukkoon kuuluu erilaisia kaavioita kuten pylväs- ja viivadiagrammeja ja tunnuslukuja. Verkkopalvelun käyttäjien seuraamiseen toteutettiin karttavisualisointi, jonka avulla voidaan seurata, mistä verkkopalvelun käyttäjät ovat kotoisin.

5.4.1 Havainnointi kaavioiden ja tunnuslukujen avulla

Erimenu.fi-verkkopalveluun toteutettu käytön seurantajärjestelmä ja seurantatiedon visualisointi koostuu pääasiassa erilaisista kaavioista. Kaaviot on toteutettu koostamalla seurantajärjestelmän keräämää tietoa. Tässä aliluvussa näytetään esimerkkejä Erimenu.fi-verkkopalvelun käyttöä havainnoivista kaavioista ja tunnusluvuista, jotka on saatu aikaa verkkopalveluun tässä työssä toteutetulla järjestelmällä.

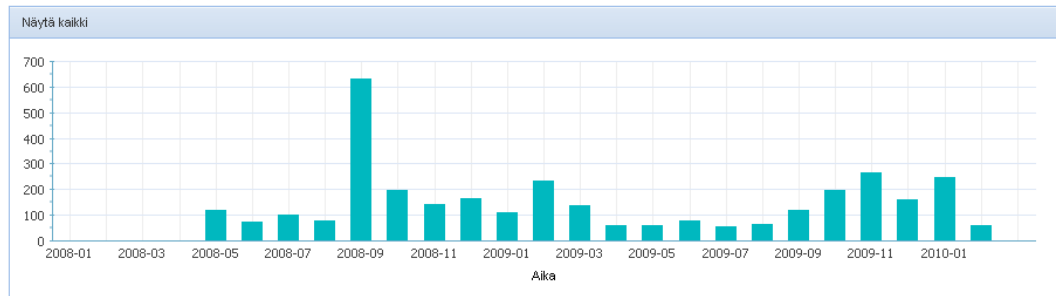
Rekisteröityneiden käyttäjien määrä on yksi verkkopalvelun käytön mittari. Rekisteröityneellä käyttäjällä on usein verkkopalvelussa laajemmat käyttöoikeudet kuin rekisteröimättömällä käyttäjällä. Verkkopalvelut toteutetaan niin, että niiden käyttö kannustaa rekisteröitymiseen laajempien ominaisuuksien avulla. Erimenu.fi-verkkopalvelussa rekisteröitynyt käyttäjä voi luoda itselleen ruokavalion, asettaa tuotteisiin avainsanoja ja lisätä ruokaohjeen omalle listalleen. Erimenu.fi:ssä oli 19.2.2010 yhteensä 3346 yksityiskäyttäjää. Käyttäjätilejä on luotu palveluun toukokuusta 2008 lähtien. Yksityiskäyttäjällä tarkoitetaan Erimenu.fi:ssä verkkopalveluun rekisteröitynyttä peruskäyttäjää. Erimenu.fi:ssä muut mahdolliset käyttäjäroolit ovat tuottaja ja ammattikäyttäjää. Kuukausittainen uusien yksityiskäyttäjien keskiarvo voidaan laskea seuraavasti:

$$\textit{käyttäjät}_{ka/kk} = \frac{\textit{käyttäjät}_{lkm}}{\textit{kuukaudet}_{lkm}} = \frac{3346}{22} \approx 150 \textit{ (käyttäjää)}.$$

Kuvasta 5.5 nähdään, että eniten uusia käyttäjätilejä on luotu syyskuussa 2008. Tämä on sama aikajakso, jolla palvelu lanseerattiin ja annettiin julkiseen käyttöön.

Yksityiskäyttäjätilien lukumäärä

Uusien yksityiskäyttäjien lukumäärä aikaväleittäin.

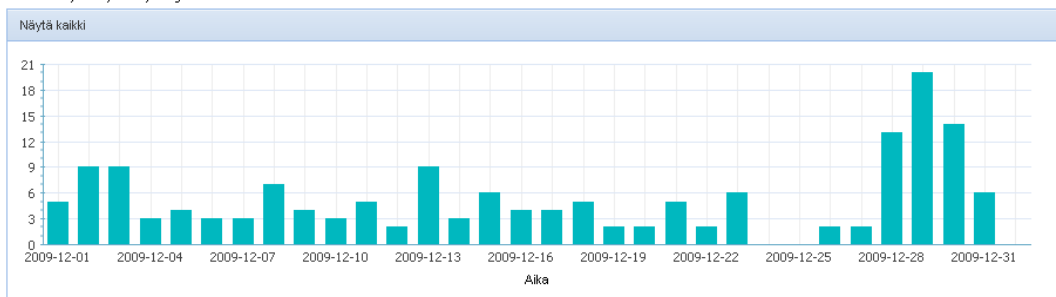


Kuva 5.5: Kuukausittainen uusien yksityiskäyttäjien määrä Erimenu.fi-verkkopalvelussa.

Kaavioiden avulla voidaan lukea muutoksia verkkopalvelun käytössä. Uusien rekisteröityneiden käyttäjien osalta voidaan tutkia esimerkiksi vuonna 2009 joulukuussa tehdyn päivityksen mahdollisia vaikutuksia. Kuvassa 5.6 nähdään, että uusien käyttäjätilien määrä on ollut samaa luokkaa joulukuun alkupuolella, mutta voidaan huomata, että välipäivinä päivityksen jälkeen uusien käyttäjätilien määrässä on selkeä piikki.

Yksityiskäyttäjätilien lukumäärä

Uusien yksityiskäyttäjien lukumäärä aikaväleittäin.

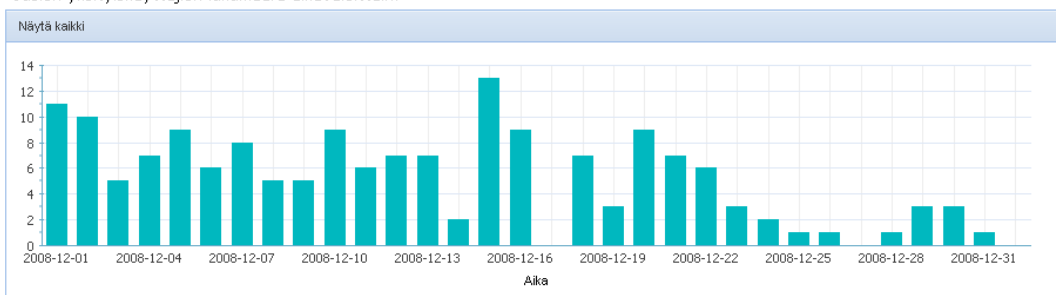


Kuva 5.6: Uusien Erimenu.fi-yksityiskäyttäjien määrä päivittäin joulukuussa 2009.

Vertailuajankohtana voidaan käyttää vuoden 2008 joulukuuta (kuva 5.7), jolloin vastaavana ajankohtana uusien käyttäjätilien määrä on alimmillaan.

Yksityiskäyttäjätilien lukumäärä

Uusien yksityiskäyttäjien lukumäärä aikaväleittäin.

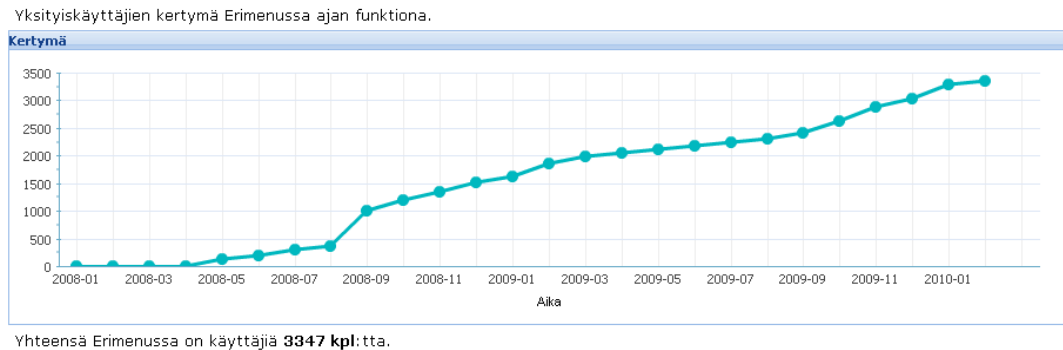


Kuva 5.7: Uusien Erimenu.fi-yksityiskäyttäjien määrä päivittäin joulukuussa 2008.

Lisääntynyt rekisteröitymismäärä voidaan havaita lisäksi tammikuun osalta. Verrattaessa vuoden 2009 tammikuun (110 uutta käyttäjää) ja vuoden 2010 tammikuun (246 uutta käyttäjää) tilastoja voidaan havaita, miten päivityksen jälkeen uusien käyttäjätilien määrä on kasvanut merkittävästi (Kuva 5.5). Prosentteina kasvu voidaan laskea seuraavasti:

$$kasvus\% = \left(\frac{käyttäjät_{1km_{joul}/2009}}{käyttäjät_{1km_{joul}/2008}} - 1 \right) * 100 = \left(\frac{246}{110} - 1 \right) * 100 \approx 120 (\%).$$

Erimenu.fi-verkkopalvelun yksityiskäyttäjien kertymästä (kuva 5.8) voidaan havainnoida, millä tavoin käyttäjien määrä on kehittynyt. Kertymäkuvaajassa peräkkäiset arvot ovat riippuvaisia toisistaan ja kuvaajaksi on valittu viivadiagrammi. Huomioitavaa kuvassa 5.8 on, että visualisointinäkyvässä kerrotaan havainnoijalle kuvaajaan liittyviä tunnuslukuja, kuten käyttäjien nykyinen kokonaismäärä.



Kuva 5.8: Yksityiskäyttäjätilien kertymä kuukausittain Erimenu.fi-verkkopalvelussa.

Erimenu.fi-verkkopalveluun toteutettiin syksyn 2009 aikana toiminnallisuus, jonka avulla palvelun käyttäjille pystyttiin lähettämään tiedotteita. Erimenu.fi:ssä tiedotteet ovat järjestelmän automaattisesti tai ylläpidon toimesta käyttäjille lähetettäviä viestejä, jotka ovat luettavissa verkkopalvelusta. Käyttäjät saavat heille lähetetyistä uusista tiedotteista tiedon heidän sähköpostiinsa. Voidaan olettaa, että koska käyttäjä saa tiedon uusista tiedotteista sähköpostiinsa, tuottaa tämä liikennettä verkkopalveluun. Eräs tiedote lähetettiin 26. tammikuuta 2010, jonka saivat kaikki verkkopalveluun rekisteröityneet käyttäjät (kuva 5.9). Erimenu.fi:ssä oli tällöin noin 3000 rekisteröitynyttä käyttäjää.

KORJAUS RUOKA-AINEESEEN AMARANTTI



Erimenu ylläpito
26.1.2010 10:30

[Merkitse lukemattomaksi](#)
[Poista viesti](#)

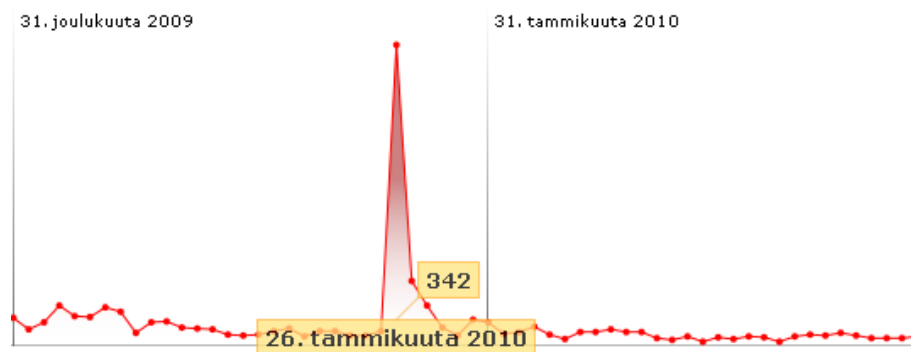
Korjaus ruoka-aineeseen amarantti

Erimenu.fi-palvelussa havaittu amaranttiin liittyvä virhe on korjattu. Ruokavalion määrittelyssä käytetyllä ruoka-ainelistalla oleva ruoka-aine **amarantti** (kiwicha tai revonhätä) **tarkoitti virheellisesti samaa kuin lisäaine E 123**. Jos esimerkiksi olet valinnut kielletyksi ruoka-aineeksi lisäaineen E 123, on kielletyksi ruoka-aineeksi valittu samalla myös amarantti sekä samaa tarkoittavat ruoka-aineet kiwicha ja revonhätä.

Kysymyksessä on kuitenkin **kaksi eri ruoka-ainetta**: amarantti(vilja) ja lisäaine E 123. Nyt Erimenu.fi-palvelusta löytyvät

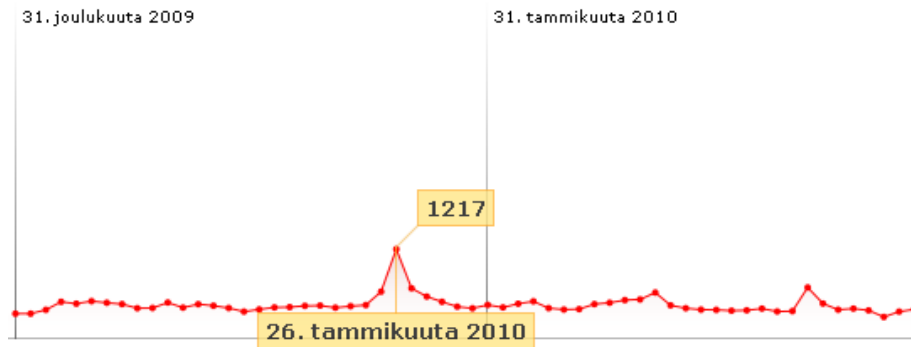
Kuva 5.9: Erimenu.fi-verkkopalvelussa luettavissa oleva tiedote.

Tiedotteen tuottama piikki havaitaan hyvin selkeästi visualisoinnissa, jossa kuvataan verkkopalveluun päivittäin kirjautuneiden käyttäjien määrää. Kyseisenä vuorokautena 26. tammikuuta 2010 Erimenu.fi-palveluun kirjautui, jopa 20 – 30 kertaa enemmän käyttäjiä kuin normaalina vuorokautena (kuva 5.10). Tiedotteen lähettämispäivänä verkkopalveluun kirjautui 342 käyttäjää kun tavallisesti kirjautuneiden käyttäjien määrä on päivittäin noin 10 – 20 luokkaa. Visualisoinnin perusteella havaitaan, että tiedote tavoitti välittömästi noin 15 %:a verkkopalvelun käyttäjistä.



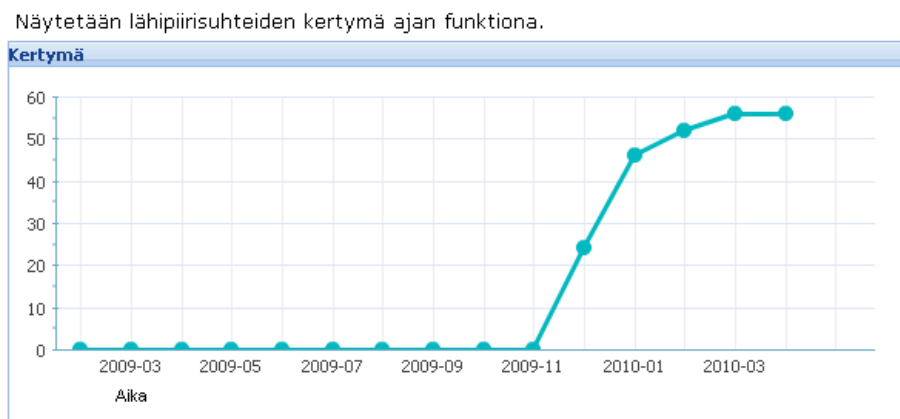
Kuva 5.10: Erimenu.fi:ssä päivittäin kirjautuneiden käyttäjien määrä.

Vastaavasti palvelun käyttöön kirjautumattomana käyttäjänä tiedotteen lähettäminen ei aiheuta vastaavanlaista piikkiä. Kuvasta 5.11 nähdään, että verkkopalvelussa vierailevien yksittäisten IP-osoitteiden määrä on kyseisenä 26. päivänä vain noin 4 – 5 kertaa suurempi kuin normaalina vuorokautena. Tiedotteen lähettämispäivänä verkkopalvelussa vieraili 1217 yksittäistä IP-osoitetta kun normaalisti verkkopalvelussa vierailee päivittäin 400 – 500 käyttäjää.



Kuva 5.11: Erimenu.fi:ssä päivittäin vierailevien IP-osoitteiden määrä.

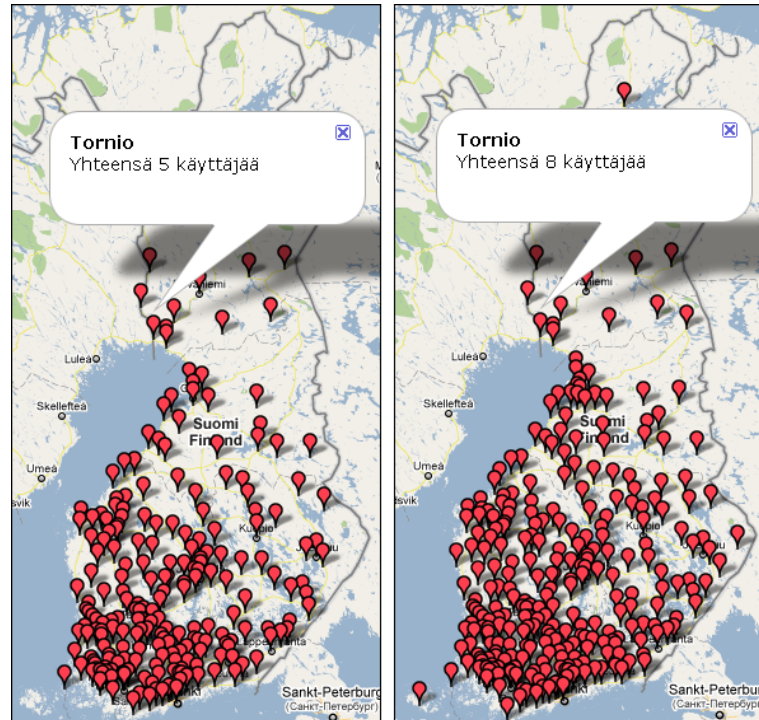
Toteutetun järjestelmän avulla pystytään kattavasti seuraamaan verkkopalveluun toteuttavien uusien ominaisuuksien käyttöönottamista. Lähipiiri-toiminnallisuus otettiin Erimenu.fi-verkkopalvelussa käyttöön marraskuussa 2009 ja kuvassa 5.12 nähdään miten Lähipiiri-toiminnallisuuden suosio on kehittynyt käyttöönoton jälkeen. Kuvasta 5.12 nähdään, että Lähipiiri-yhteyksiä luotiin eniten heti toiminnallisuuden julkistamisen jälkeen, jonka jälkeen uusien Lähipiiri-yhteyksien määrä pienentynyt.



Kuva 5.12: Erimenu.fi-verkkopalvelussa tehtyjen Lähipiiri-yhteyksien kertymä.

5.4.2 Havainnointi karttavisualisointien avulla

Erimenu.fi-verkkopalvelussa käyttäjien pakollisiin käyttäjätietoihin kuuluu käyttäjän paikkakunta. Paikkakunta on ennalta määrätystä listasta valittava tieto, joka tallennetaan käyttäjäkohtaisesti tietokantaan. Tähän tietoon verkkopalvelusta voidaan tuottaa karttavisualisointi, joka kuvaa Erimenu.fi-verkkopalvelun laajentumista uusille paikkakunnille. Kuvassa 5.13 nähdään esimerkkivisualisointi Erimenu.fi-verkkopalveluun rekisteröityneistä käyttäjistä paikkakunnittain vuoteen 2009 ja vuoteen 2010 mennessä.



Kuva 5.13: Vasemmalla kuvataan Erimenu.fi-palveluun rekisteröityneet käyttäjät vuoteen 2009 mennessä ja oikealla vuoteen 2010 mennessä.

Käytön seurannan visualisointityökaluun toteutettiin mahdollisuus hakea verkkopalvelun käyttäjien paikkakunnat kuukauden tarkkuudella. Analysoijan on mahdollista valita haluamansa aikaväli tarkastelujaksoksi. Tällöin voidaan tutkia onko tapahtumilla, kuten tietyllä paikkakunnalla järjestetyillä messuilla, joissa verkkopalvelua esitellään, vaikutusta siltä alueelta rekisteröityneiden käyttäjien määrään.

Karttavisualisointia muokattiin Google Maps -palvelun⁵⁸ tarjoamaan toteutukseen verrattuna merkitsijöiden (marker) osalta niin, että merkitsijöiksi valittiin ole-tusta pienemmät merkit. Näin tiheästäkin olevat merkit eivät asetu toistensa päälle ja visualisointi on selkeämpi kuin suuremmilla merkitsijöillä. Kuvassa 5.13 nähtävää visualisointia voitaisiin monipuolistaa hyödyntämällä lämpötehostetta (lämpökartta, heat map) kaupunkien merkitsemässä, jolloin enemmän käyttäjiä sisältävät kaupungit merkittäisiin voimakkaammin kuin vähemmän käyttäjiä sisältävät. Näin kartasta voitaisiin helpommin lukea ja vertailla kaupunkien välisiä painoarvoja ja muutosta niissä.

5.5 Muutoksen selittäminen

Muutoksen selittäminen eli käytön seurantatiedon analysointi rajataan lähtökoh-taisesti tämän työn ulkopuolelle. Havaitun käytön seurantatiedon avulla ei voida autenttisesti sanoa, että havaittu muutos johtuu tietystä yksittäisestä tekijästä.

⁵⁸<http://maps.google.com/>

Esimerkiksi tähän työn indikaattoriksi toteutetulla Lähipiiri-ominaisuudella pyrittiin luomaan verkkopalvelun käyttöön muutosta, joka pystyttäisiin havaitsemaan. Vaikka muutos havaitaan, on mahdollista, että se johtuu täysin muista tekijöistä tai tekijöiden yhteisvaikutuksesta.

Havaittua muutosta voidaan yrittää selittää tiedossa olevilla tekijöillä, mutta syy-seuraus -yhteyksien absoluuttinen määrittäminen vaatii käyttäjien osallistumisen käytön seurantatiedon keräämiseen. Käyttäjien osallistuminen voidaan toteuttaa esimerkiksi käyttäjäkyselyin, jolloin käyttäjät voivat kertoa, mistä syystä he alkoivat käyttää verkkopalvelua tai sen osaa aiempaa enemmän. Tämän työn puitteissa tämänkaltainen käyttäjäkysely ei ole laajuutensa vuoksi mahdollinen. Käyttäjäkyselyn suunnittelu ja toteuttaminen on kuitenkin tämän työn perusteella mahdollista.

6 Tulosten arviointi

Diplomityössä toteutettua Erimenu Logger käytön seurantajärjestelmää arvioidaan kahdesta näkökulmasta. Tulosten arviointi on tärkeä osa diplomityötä, koska siinä verrataan työssä saatuja tuloksia aihealueen muuhun tutkimukseen ja olemassa oleviin järjestelmiin. Ensinnäkin verrataan miten sovellustasolla toimiva käytön seurantajärjestelmällä on mahdollista kerätä tietoa suhteessa palvelinpäässä ja asiakasohjelmissa toimiviin järjestelmiin. Toiseksi Erimenu Loggeria verrataan Erimenu.fi-verkkopalvelussa käytössä oleviin käytön seurantajärjestelmiin Google Analytics ja AWStats. (Vrt. Järvinen & Järvinen, 2004, ss. 35-36)

Verkkopalvelujen käytön seurantaan liittyy hyvin monia arviointiin vaikuttavia tekijöitä. Monikriteerimenetelmät (multi-criteria decision analysis (MCDA) tai multi-criteria decision making (MCDM)) tarjoavat systemaattisen lähestymistavan yhdistää monentasoista tietoa päätöksenteon tukemiseksi. MCDA-mallien avulla pystytään yhtäaikaiseksi keskenään vertailemaan ominaisuuksia ja teknologioita joilla ominaisuudet on toteutettu. MCDA-mallien avulla on mahdollista arvioida tuloksia kun arvioitavina ovat ominaisuudet ovat ristiriidassa keskenään. MCDA-malli osoittaa vaihtoehtojen väliset ristiriidat ja luo pohjan tulosten perusteella läpinäkyvästi tehtävälle päätökselle. MCDA-malli ei pyri esittämään yksikäsitteistä ratkaisua ongelmaan vaan sen avulla luodun arvioinnin perusteella voidaan tehdä eri johtopäätöksiä riippuen arvioinnin kohteesta ja arvioinnin tekevästä henkilöstä. MCDA-mallin avulla pyritään esittämään arvioinnin kohteeseen liittyvät sidokset mahdollisimman subjektiivisesti.

Diplomityön tuloksia arvioidaan MCDA-mallin mukaisella MAGIQ-menetelmällä (Multi-Attribute Global Inference of Quality). MAGIQ on analyttinen hierarkiaprosessi -menetelmä. Analyttiset hierarkiaprosessimenetelmät ovat yleisimmin käytettyjä MCDA-menetelmiä ja niiden avulla on mahdollista tehdä optimoituja ratkaisuja. MAGIQ-menetelmässä vaihtoehdot ja niiden ominaisuudet asetetaan taulukkoon, jossa niille annetaan numeroarvo. Tässä työssä arviointiin käytetään yksinkertaista "0" (ei ole mahdollista toteuttaa) tai "1" (on mahdollista toteuttaa) arviointia sen mukaan onko ominaisuus mahdollista toteuttaa kyseisen teknologian avulla. MAGIQ-menetelmässä on mahdollista antaa ominaisuuksille painoarvot, joiden avulla haluttujen ominaisuuksien merkitystä voidaan korostaa tai pienentää. Tämän ei ole nähty olevan tarpeellista tämän työn arvioinnissa, koska tarkoituksena on arvioinnin avulla tehdä näkyväksi miten teknologiat toimivat toisiinsa nähden, eikä yksikäsitteisesti näyttää mikä teknologioista on paras. (Kiker et al., 2005, ss. 95-108; vrt. McCaffrey, 2009, ss. 738-742; vrt. Verta, 2006, ss. 39-41)

Seurantateknologioiden osalta vertailu pohjautuu Croll & Sean:n määrittelemiin palvelinpään käytön seuraimen ja asiakasohjelmassa toimivan seuraimen hyviin

ja huonoihin puoliin. Palvelimella verkkopalvelun käyttöä seuraavan järjestelmän hyviä puolia ovat kerätyn seurantatiedon yksityisyys, riippumattomuus käyttäjän asiakasohjelmasta ja ettei seurantatiedon kirjaaminen vaikuta verkkopalvelun sivujen latausaikaan. Palvelinpäessä toimivan seuraimen heikkouksia ovat, että seurantatieto ei ole aina verkkopalvelun kehittäjän saavutettavissa ja ettei palvelinpään keräimen näkevä seurantatieto ole kattavaa. Asiakasohjelmassa toimivalla järjestelmällä pystytään palvelinpään seurantajärjestelmää laajemmin keräämään käyttäjään liittyvää tietoa, koska seurantajärjestelmä näkee kaiken minkä verkkopalvelun käyttäjäkin näkee. Asiakasohjelmassa toimiva käytön seurantajärjestelmä pääsee käsiksi asiakasohjelman tietoihin kuten viimeiseen vierailupäivään ja resoluutiotietoihin. Asiakasohjelmassa toimivan käytön seurantajärjestelmän huonoja puolia ovat riippuvaisuus käyttäjän asiakasohjelmasta ja että seuranta aiheuttaa verkkopalvelujen sivujen lataamisen hidastumista. (Vrt. Croll & Sean, 2009, ss. 114-129)

Croll & Sean eivät ota kantaa sovellustason käytön seurantajärjestelmään. Sovellustason käytön seurantajärjestelmän etu suhteessa palvelinpään ja asiakasohjelmassa toimivaan järjestelmään on, että sovellustason seuraimella on pääsy verkkopalvelun tietosisältöihin. Sovellustasolle sisällytettävän seurantajärjestelmän merkittävimmät heikkoudet liittyvät teknologian heikkoon siirrettävyyteen verkkopalvelujen välillä. Heikko siirrettävyys tarkoittaa, että järjestelmän toteutus ja käyttöönotto ovat haastavampia kuin palvelinpään seuraimessa ja asiakasohjelmassa toimivassa seuraimessa.

Taulukossa 6.1 kerätään yhteen teknologioiden mahdollistamat käytön seurantaominaisuudet. Taulukko 6.1 on tehty mukaillen Croll & Sean, s. 129 määrittelemää teknologioiden vertailutaulukkoa, mutta lisäten sovellustason käytön seuraimelle ominaisia kohtia. Taulukosta on jätetty pois verkkopalvelujen monitorointiin liittyviä kohtia, koska ne eivät ole oleellisia tämän diplomityön kannalta, koska tarkoituksena on tuoda esiin verkkopalvelun käyttäjien toimia.

Taulukko 6.1: Käytön seurantatekniikoiden vertailutaulukko.

	Palvelinseuranta	Asiakasseuranta	Sovellusseuranta
Käyttöönotto			
Tiedon yksityisyys	Kyllä	Kyllä / Ei	Kyllä
Asiakasohjelma-riippumaton	Kyllä	Ei	Kyllä
Lisää koodia sovellukseen	Ei	Kyllä	Kyllä
Tiedon näkyvyys			
POST-arvot	Ei	Kyllä	Kyllä
Asiakasohjelman sisäiset tiedot	Ei	Kyllä	Kyllä
Verkkopalvelun tietosisällöt	Ei	Ei	Kyllä
Näkee 404-virheet	Kyllä	Ei	Ei
Näkymien nimeäminen	Ei	Kyllä / Ei	Kyllä
Suorituskyky			
Lisää palvelin-kuormaa	Kyllä	Ei	Kyllä
Lisää sivulataus-kuormaa	Ei	Kyllä	Kyllä

Erimenu.fi-verkkopalvelussa on käytössä Erimenu Loggerin myötä kolme erillistä käytön seurantatekniikkaa ja niihin perustuvaa järjestelmää. Google Analytics-palvelu seuraa verkkopalvelun käyttöä asiakasohjelmassa, AWStats-järjestelmä koostaa näkymiä perustuen palvelintasolla kerättävään lokitietoon ja diplomityössä toteutettu Erimenu Logger toteuttaa seurannan sovellustasolla. Taulukossa 6.2 esitetään kuinka monipuolista tietoa järjestelmät tuottavat. Taulukon 6.2 kohdat on valittu mukailen Croll & Sean, s. 129 esittämää vertailutaulukkoa.

Taulukko 6.2: Käytön seurantajärjestelmien vertailutaulukko.

	AWStats	Google Analytics	Erimenu Logger
Käyttöönotto			
Tiedon yksityisyys	Kyllä	Ei	Kyllä
Asiakasohjelma-riippumaton	Kyllä	Ei	Kyllä
Lisää koodia sovellukseen	Ei	Kyllä	Kyllä
Tiedon näkyvyys			
POST-arvot	Ei	Ei	Kyllä
Asiakasohjelman sisäiset tiedot	Ei	Kyllä	Ei
Verkkopalvelun tietosisällöt	Ei	Ei	Kyllä
Näkee 404-virheet	Kyllä	Ei	Ei
Näkymien nimeäminen	Ei	Ei	Kyllä
Suorituskyky			
Lisää palvelin-kuormaa	Kyllä	Ei	Kyllä
Lisää sivulataus-kuormaa	Ei	Kyllä	Kyllä

6.1 Verrattuna palvelinpään käytön seuraimeen

Palvelinpäässä toimiva käytön seurantajärjestelmä pääsee käsiksi vain HTTP-palvelimelle välitettyyn tietoon. Tähän tietoon kuuluvat pääasiassa tiedot, jotka ovat pakollisia verkkopalvelun käytön kannalta. Palvelinpäässä toimivalla käytön seurantajärjestelmällä ei ole siis pääsyä verkkopalvelun tietosisältöihin tai verkkopalvelun käyttäjän asiakasohjelman ympäristömuuttujiin. Verkkopalvelun yhteyteen toteutettu ja siten verkkopalvelun kaikkiin osiin käsiksi pääsevä käytön seurantajärjestelmä pääsee käsiksi laajemmin verkkopalvelun osiin kuin seurain, jolla on pääsy vain palvelimen käsittelemiin tietoihin. Toisaalta palvelimen keräämistä tiedoista on helpompaa kerätä tiettyä tietoa, kuin olisi erillisen keräin sovelluksen toteuttaminen sovelluksen yhteyteen. Esimerkiksi käytetyn kaistan määrän seuranta on suoraan palvelimen ja palvelimen lokitiedoista helpompaa kuin, että tämä tiedon keräämiseen toteutettaisiin omat työkalut, jotka toimisivat sovelluksen yhteydessä.

Toteutettu sovellustasolle integroitava käytön seurantajärjestelmä mahdollistaa verkkopalvelussa tapahtumien muutosten seuraamisen laajemmin kuin palvelinpäässä toimiva käytön seurantajärjestelmä. Aliluvussa 5.4 esitetty amarantti-päivitys näkyy Apache-palvelimen tuottamaa lokitietoa visualisoivassa AWStats-ohjelmassa kuvassa 6.1 esitetyllä tavalla.

Päivä	Vierailujen määrä	Sivuja	Osumia	Kaista
16.01.2010	367	1656	16362	100.88 Mt
17.01.2010	399	1989	18636	138.27 Mt
18.01.2010	415	2659	20181	164.83 Mt
19.01.2010	430	2426	20768	119.00 Mt
20.01.2010	429	2552	18948	124.49 Mt
21.01.2010	426	2933	21582	201.98 Mt
22.01.2010	424	3038	22011	138.24 Mt
23.01.2010	433	2414	21109	133.71 Mt
24.01.2010	451	1959	19887	129.93 Mt
25.01.2010	647	2643	30174	167.46 Mt
26.01.2010	1280	9129	67280	341.17 Mt
27.01.2010	713	3942	31860	183.33 Mt
28.01.2010	564	3196	26987	149.70 Mt
29.01.2010	491	2048	20718	122.65 Mt

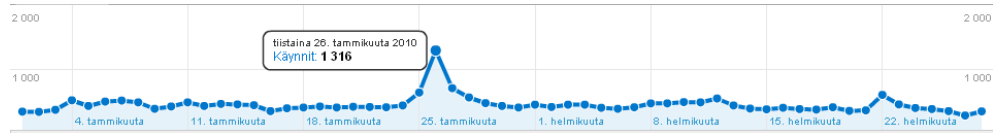
Kuva 6.1: Erimenu.fi-palvelussa tehty amarantti-päivitys näkyy selkeästi HTTP-palvelimen lokitietoa jäsentävässä AWStats-ohjelmassa.

Kuvassa 6.1 nähtävä yksittäisten käyttäjien määrä 26.1.2010 on 1280 käyttäjää, joka vastaa Erimenu Logger -järjestelmällä saatua arvoa 1217 käyttäjää. Voidaan sanoa, että toteutettu järjestelmä toimii, koska sillä kerätty ja koostettu käytön seurantatieto on yhtä yleisesti tunnetun AWStats-järjestelmän kanssa.

6.2 Verrattuna asiakasohjelmassa toimivaan käytön seuraimen

Asiakasohjelmassa toimivalla käytön seurantajärjestelmällä on useita etuja verrattuna palvelinpäässä toimivaan käytön seuraimen, koska käyttäjän asiakasohjelmassa käytön seuranta mahdollistaa asiakasohjelmaan liittyvien ympäristömuuttujien keräämisen. Asiakasohjelmassa toimivalla käytön seurantajärjestelmällä on lisäksi käytössään merkittävimmät HTTP-palvelimen keräämät tiedot. Voidaan sanoa, että asiakasohjelmassa toimiva käytön seurantajärjestelmä tuottaa käytön seurantatietoa palvelinpään seurainta laajemmin.

Sovellustasolla toimivaan käytön seurantajärjestelmään verrattuna asiakasohjelmassa suoritettavalla käytön seurantajärjestelmällä on samat heikkoudet kuin palvelintasolla toteutettavassa seurannassa. Asiakasohjelmassa toimivalla käytön seurantajärjestelmällä ei ole pääsyä verkkopalvelun tietosisältöihin. Kuvassa 6.2 nähdään miten aliluvussa 5.4 kuvattu amarantti-päivitys havaitaan asiakasohjelmatasolla toimivassa Google Analytics -käytön seurantatiedon keräyspalvelussa.



Kuva 6.2: Päivittäisten käyttäjien määrä Erimenu.fi-verkkopalvelussa aikajaksolla tammi-helmikuu 2010 Google Analytics -palvelussa nähtynä.

Kuvassa 5.4 nähdään verkkopalvelussa vierailleiden yksittäisten käyttäjien määrä. Amarantti-päivityksen päivänä verkkopalvelussa vieraili Google Analyticsin mukaan 1316 käyttäjää, joka vastaa Erimenu Loggerin rekisteröimää 1217 käyttäjää. Lukujen välinen erotus voidaan selittää seurantajärjestelmien erilaisella käyttäjien suodattamis- ja rekisteröimismenetelmillä. On mahdollista esimerkiksi, että Google Analytics tulkitsee samasta IP-osoitteesta samana päivänä tulevat sivulatauspyynnöt eri käyttäjiksi jos käytön välillä on pitkä aika. Erimenu Loggerissa lasketaan suoraan verkkopalvelussa päivittäin vierailevien yksittäisten IP-osoitteiden määrää.

7 Yhteenveto

Verkkopalvelujen käytön seuranta on ajankohtainen tutkimusalue ja työssä aihealuetta käsiteltiin laajasti. Aliluvussa 1.1 esitettiin tutkimuskysymyksiin työssä pystyttiin vastaamaan kattavasti. Työssä selvitettiin millä teknologiaratkaisuilla verkkopalveluiden seuraaminen on mahdollista. Havaittiin, että toteutusteknisesti käytön seuranta voidaan jakaa palvelintason käytön seurantaan, asiakasohjelmassa tapahtuvaan käytön seurantaan ja sovellustason seurantaan. Palvelintasolla seuranta tapahtuu HTTP-palvelimella. Palvelintasolla tapahtuvalle seurannalle on ominaista, että se keskittyy verkkopalveluun teknisestä näkökulmasta. Palvelintasolla käytön seuranta pohjautuu HTTP-palvelimen tallentamien HTTP-protokollaan liittyvien tietojen keräämiseen. Käyttäjän asiakasohjelmassa suoritettava käytön seuranta pääsee palvelintason laajemmin käsiksi käyttäjään liittyviin tietoihin. Asiakasohjelmaan liittyvän seurantajärjestelmän heikkous on, riippuvaisuus asiakasohjelmasta, joka tarkoittaa, että käyttäjän on halutessaan mahdollista kiertää seuranta. Sovellustason seurannalla tarkoitetaan suoraan verkkopalvelun muiden osien yhteyden sisällytettävää järjestelmää, jolla on pääsy verkkopalvelun tietosisältöihin ja ajon-aikaisiin muuttujiin.

Diplomityössä toteutettiin sovellustasolle integroitava käytön seurantajärjestelmä Erimenu.fi-verkkopalveluun. Tietojen kerääminen jaettiin kahteen loogiseen kokonaisuuteen: sivulataustapahtumien seurantaan ja tietosisältöjen seurantaan. Sivulatausten seuranta on monin osin rinnakkainen palvelintasonseuraimen ja asiakasohjelmassa toimivan seuraimen kanssa, mutta sovellustasolla sivulataustapahtumiin on mahdollista liittää tieto verkkopalvelukohtaisista tiedoista. Tietosisältöjen osalta sovellustasolla toimiva käytön seuranta tuo palvelinpään- ja asiakasohjelmassa toimiviin käytön seurantajärjestelmiin verrattuna täysin uuden näkökulman, koska niillä ei ole pääsyä verkkopalveluiden tietojärjestelmiin. Käytön seurannan toteuttavan teknologian valinta vaikuttaa mitä tietoa verkkopalvelun käytöstä voidaan kerätä. Esitetyt asiat vastaavat työn alussa esitettyyn tutkimuskysymykseen “millä tavoin toteutettavan seurantajärjestelmä keräämä tieto on kattavampaa kuin vaihtoehtoisilla teknologioilla ja järjestelmillä kerättävä tieto?”.

Työssä havaittiin, että pelkkä käytön seurantatiedon kerääminen ei ole riittävää vaan kerättyä tietoa täytyy koostaa ja visualisoida, jotta verkkopalvelun käyttöä voidaan havainnoida. Työssä todettiin, että käytön seurannan visualisointi on hyvä toteuttaa suoraan visualisoitavan järjestelmän yhteyteen, koska tämä parantaa tiedon saavutettavuutta. Työssä kuvattiin mitä käytön seurantatiedon visualisoinnissa tulee ottaa huomioon ja minkälaisilla teknologioilla visualisointi on mahdollista tehdä. Esitetyt ratkaisut vastaavat tutkimuskysymykseen: “miten kerätty seuranta-tieto esitetään, että sen havaitseminen on mahdollisimman havainnollista?”.

Erimenu.fi-verkkopalveluun toteutetun käytön seurantajärjestelmän avulla on mahdollista kerätä verkkopalvelun käyttöön liittyvää kvantitatiivista tietoa. Tieto kerätään automaattisesti ilman käyttäjän väliintuloa. Kerätyn tiedon pohjalta toteutettiin verkkopalvelun yhteyteen kattavat visualisointinäkymät, joiden avulla ylläpitokäyttäjän on mahdollista seurata miten verkkopalvelua käytetään. Toteutus verkkopalvelun yhteyteen parantaa käytön seurantatiedon saavutettavuutta, koska tiedon lähteenä on tiedon lukijalle luonnollinen ympäristö. Toteutetut visualisoinnit koostuvat erilaisista kuvaajista, kartoista ja tunnusluvuista. Visualisoinnit toteutettiin kattaviksi ja kuvaajat toteutettiin niin, että havainnoitsijalla on vapaus vaikuttaa mitä tietoa ja miltä aikaväliltä kuvaajissa esitetään. Käytön seurantajärjestelmän keräämän tiedon perusteella on jatkossa mahdollista toteuttaa tarpeen mukaan uusia visualisointeja. Toteutetulla seurantajärjestelmällä on mahdollista siten seurata verkkopalveluun toteuttavien uusien ominaisuuksien käyttöönottoa ja toisaalta tutkia uusien ominaisuuksien vaikutusta palvelun käyttöön muiden ominaisuuksien osalta.

Työn tulosten perusteella osoitettiin, että sovellustasolla toimivan käytön seurantajärjestelmän avulla pystytään seuraamaan laajemmin verkkopalvelun tietosisältöihin liittyvää tietoa. Erimenu.fi-verkkopalvelussa aikaisemmin käytössä olleilla AWStats ja Google Analytics -seurantajärjestelmillä verkkopalvelun tietosisältöihin ei päästä kiinni. Tutkimuksessa osoitettiin, että valittujen ominaisuuksien osalta toteutettu järjestelmää lisäksi kerää ja sen avulla on mahdollista saada näkyville sama tieto kuin mitä edellä mainituilla järjestelmillä. AWStatsin ja Google Analyticsin kaltaisten järjestelmien etuina ovat helppo asennettavuus ja ettei järjestelmää tarvitse ylläpitää. Sovellustasolle integroituvalla järjestelmälle täytyy aina esitellä verkkopalvelukohtaiset tietosisällöt ja miten verkkopalvelun ominaisuuksien toiminta, jotta seurantajärjestelmän pystyy tallentaa niihin liittyvän tiedon. Kustannuksiltaan sovellustason järjestelmä on kalliimpi, mutta tuottaa enemmän arvokasta tietoa.

Toteutettu järjestelmä on pääasiallisesti tarkoitettu Erimenu.fi-verkkopalvelun ylläpitäjien työkaluksi, jolla he voivat tarkkailla verkkopalvelun käyttöä. Työssä todettiin, että seurantatiedosta toteutettuja visualisointeja on mahdollista tuoda myös tavallisten käyttäjien saataville ja näin parantaa verkkopalvelun käyttökokemusta. Yksittäisissä tapauksissa käytön seurantatietoa tuotiin käyttäjien saataville tämän työn puitteissa, mutta pääasiallisesti seurantatiedon tuominen tavallisten käyttäjien saataville jätettiin jatkokehityksen varaan.

Kriittisestä näkökulmasta työn sisältämä kokonaisuus, joka kattaa käytön seurantajärjestelmän toteutuksen sekä kerättävän tiedon visualisoinnin, oli liian laaja käsiteltäväksi yhdessä diplomityössä. Työssä oltaisiin päästy syvemmälle kun toteutuksen osalta oltaisiin voitu keskittyä vain tiedon keräämiseen tai tiedon

esittämiseen. Koko prosessin kuvaavana työ jää tietyiltä osilta pintapuoliseksi eikä ongelmakohtiin pystytty keskittymään paremmin. Toteutetun järjestelmän yleistettävyys verkkopalvelu riippumattomasti on yksi toiminnallisuus, jonka kehittämiseen olisi voitu keskittyä jos työ olisi rajattu paremmin. Työn sisältö ja toteutettavana oleva järjestelmä oli riippuvainen projektissa annetusta tehtävästä, joten kaikilta osin keskittyminen vain tiettyyn osaan verkkopalvelujen käytön seurannassa ei kuitenkaan olisi ollut mahdollista. Käytön seurantatiedon kerääminen ja kerätyn tiedon visualisointi ovat liitoksissa toisiinsa eikä voida nähdä, että toista voitaisiin toteuttaa ilman toista. Jos työssä olisi keskitytty pelkästään seurantatiedon visualisointiin perustuisi työ hypoteeseihin eikä aitoon käytön seurantatietoon. Toisaalta jos työssä olisi toteutettu vain käytön seurantatietoa keräävä järjestelmä olisi työn tulosten arviointi vaikeaa. Mahdollisesti optimaalinen tilanne olisi ollut jos projektiin liittyen olisi toteutettu rinnakkain kaksi erillistä diplomi- tai muuta tutkimustyötä, joissa olisi keskitytty erityisesti näiden kahden osakokonaisuuden (tiedon kerääminen ja tiedon visualisointi) toteutukseen.

Aikaisempaan tutkimukseen ((Silius et al., 2009) ja (Silius et al., 2010)) verrattuna työssä tutkittiin etenkin millä tavoilla verkkopalvelujen käytön seuraaminen on mahdollista toteuttaa ja miten verkkopalveluun voidaan toteuttaa sovellustasolla verkkopalvelun käyttöä seuraava järjestelmä. Aikaisemmassa tutkimuksessa on keskitytty verkkopalveluiden prosesseista saatavan tiedon esittämiseen matemaattisesti formaalilla tavalla, kun nyt tarkoituksena oli tuottaa etenkin verkkopalvelun ylläpitäjille tarkoitettu järjestelmä, jonka avulla verkkopalvelun käytöstä voidaan kerätä ja esittää automatisoidusti kaikki mahdollisten kvantitatiivinen seurantatieto.

Tulevaisuudessa toteutetun järjestelmän avulla on mahdollista seurata Erime-nu.fi-verkkopalvelun käyttöä ja hyödyntää tutkimuksessa eteen tulleita kokemuksia tulevissa projekteissa.

Lähteet

- Ajax widget frameworks survey*. (2009, 6. elokuu). Hypermedian laboratorio.
Lainattu 2009, 16. marraskuuta, saatavilla http://matriisi.ee.tut.fi/wiki/index.php?title=Ajax_Widget_Frameworks_Survey&oldid=20587 (Vaatii erilliset tunnukset)
- Alanto, H., Artte, U., Huhtala, H., Karonen, P., Koskiniemi, T., Kosunen, R., et al. (2010). *Sosiaalisen median sanasto*. Albertinkatu 23 A 1, 00120 Helsinki: Sanastokeskus TSK ry. Saatavilla http://www.tsk.fi/tiedostot/pdf/Sosiaalisen_medan_sanasto (ISBN: 978-952-9794-26-3)
- Antill, J. (2007, toukokuu). *Hypertext transfer protocol*. And.org. Lainattu 2009, 16 marraskuuta, saatavilla <http://www.and.org/texts/server-http>
- Berners-Lee, T., Bray, T., Connolly, D., Cotton, P., Fielding, R., Jeckle, M., et al. (2004, 15. joulukuu). *Architecture of the world wide web, volume one*. W3C. Lainattu 2010, 23. helmikuuta, saatavilla <http://www.w3.org/TR/webarch/>
- The biggest google analytics sites*. (2009). Backend Battles. Lainattu 2010, 16. helmikuuta, saatavilla http://www.backendbattles.com/backend/Google_Analytics
- Blaauw, G. A., Brooks, F. P., & Buchholz, W. (1962). *Planning a computer system; project stretch* (W. Buchholz, toim.). McGraw-Hill Book Company, Inc. Lainattu 2010, 22. maaliskuuta, saatavilla http://archive.computerhistory.org/resources/text/IBM/Stretch/pdfs/Buchholz_102636426.pdf (Chapter 7, Variable-Field-Length Operation)
- Browser statistics*. (2009). W3C. Lainattu 2009, 13. marraskuuta, saatavilla http://www.w3schools.com/browsers/browsers_stats.asp
- Burbeck, S. (1987). *Applications programming in smalltalk-80(tm): How to use model-view-controller (mvc)*. ParcPlace Systems, Inc. Lainattu 2010, 25. tammikuuta, saatavilla <http://st-www.cs.illinois.edu/users/smarch/st-docs/mvc.html>
- Bustos-Jimenez, J., Caromel, D., Costanz, A., & Piquer, J. M. (2005, marraskuu). Balancing active objects on a peer-to-peer infrastructure. XXV International Conference of the Chilean Computer Science Society (SCCC'05). Lainattu 2009, 13. marraskuuta, saatavilla <http://mario.leyton.cl/publications/sccc05.pdf> (ISBN: 0-7695-2491-5)

- Croll, A., & Sean, P. (2009). *Complete web monitoring* (1. p.; S. Laurent, toim.). O'Reilly Media, Inc. (ISBN:9780596155131)
- Dickson, P. R., & Ginter, J. L. (1987, huhtikuu). Market segmentation, product differentiation, and marketing strategy. , 10. *The Journal of Marketing*, Vol. 51, No. 2. Saatavilla <http://www.jstor.org/stable/1251125>
- Distributed application architecture*. (2007). Sun Microsystem. Lainattu 2009, 13. marraskuuta, saatavilla
<http://java.sun.com/developer/Books/jdbc/ch07.pdf>
- Everi, A. (2006). *Lokitietoanalyysi verkkopalvelun mallintamisen ja analysoinnin välineenä* [Diplomityö]. Saatavilla
http://matriisi.ee.tut.fi/hypermedia/julkaisut/di_antti_everi.pdf
- Ext js; charts*. (s.a.). Ext JS. Lainattu 2010, 23. helmikuuta, saatavilla
<http://www.extjs.com/depoy/dev/examples/chart/charts.html>
- Fallside, D., & Walmsley, P. (toim.). (2004, 28. lokakuu). *Xml primer*. W3C. Lainattu 2010, 23. helmikuuta, saatavilla
<http://www.w3.org/TR/xmlschema-0/>
- Garshol, L. M. (1999). *How the web works: Http and cgi explained*. Lainattu 2009, 16. marraskuuta, saatavilla
<http://www.garshol.priv.no/download/text/http-tut.html>
- Garson, D. (2008, 8. marraskuu). *Case studies*. North Carolina State University. Lainattu 2010, 9. helmikuuta, saatavilla
<http://faculty.chass.ncsu.edu/garson/PA765/cases.htm>
- German court says ip addresses in server logs are not personal data*. (2008, 14. 10). Out-Law. Lainattu 2010, 9. helmikuuta, saatavilla
<http://www.out-law.com/page-9505>
- Google analytics*. (2010, 12. helmikuu). Wikipedia. Lainattu 2010, 16. helmikuuta, saatavilla http://en.wikipedia.org/w/index.php?title=Google_Analytics&oldid=343518418
- Graham, P. (2005, marraskuu). *Web 2.0*. Lainattu 2010, 7. maaliskuuta, saatavilla
<http://www.paulgraham.com/web20.html>
- Grönfors, M. (1982). *Laadullisen tutkimuksen kenttätömenetelmät* (1. p.). WSOY. Lainattu 2009, 16. marraskuuta, saatavilla
http://homepage.mac.com/vilkka/Kirjat/Laadullisen_tutkimuksen.pdf
(ISBN:9789512657629)

- Gube, J. (2009, 2. tammikuu). *10 promising free web analytics tools*. Six Revisions. Lainattu 2010, 16. helmikuuta, saatavilla <http://sixrevisions.com/usabilityaccessibility/10-promising-free-web-analytics-tools/>
- Haikala, I., & Märijärvi, J. (2002). *Ohjelmistotuotanto* (8. p.). Talentum Media Oy. (Alkuaan julk. 1995; ISBN:9789521404863)
- Hallam-Bake, P. M., & Behlendorf, B. (s.a.). *Extended log file format*. Lainattu 2010, 15. maaliskuuta, saatavilla <http://www.w3.org/TR/WD-logfile-960221.html>
- Henkilötietolaki*. (1999, 22. huhtikuu). FINLEX Valtion säädöstietopankki. Lainattu 2010, 24. helmikuuta, saatavilla <http://www.finlex.fi/fi/laki/ajantasa/1999/19990523>
- Henry, A. (2007, 27. elokuu). *10 free, innovative web analytics tools*. ReadWriteWeb. Lainattu 2010, 16. helmikuuta, saatavilla http://www.readwriteweb.com/archives/10_web_analytics_tools_free_innovative.php
- Hieta-Wilkman, S. (s.a.). *Tiedon hallinta; mitä tiedon/tietämyksen hallinta on?* Helsingin yliopisto. Lainattu 2010, 24. helmikuuta, saatavilla <http://www.valt.helsinki.fi/comm/areena/comm/tieto.htm>
- Hiltunen, A. (1999). *MuotINVALMISTUKSEN TUOTETIEDON HALLINTA* [Diplomityö]. Otamedia Oy. Saatavilla <http://www.tkk.fi/Units/Production/Publications/tkk-kpt-3-99.PDF> (ISBN: 951-22-5712-2)
- How apache came to be*. (2009). Apache. Lainattu 2009, 24. marraskuuta, saatavilla http://httpd.apache.org/ABOUT_APACHE.html
- How dns works*. (2003, maaliskuu). Microsoft. Lainattu 2009, 16. marraskuuta, saatavilla [http://technet.microsoft.com/en-us/library/cc772774\(WS.10\).aspx](http://technet.microsoft.com/en-us/library/cc772774(WS.10).aspx)
- Http - hypertext transfer protocol*. (2009). W3C. Lainattu 2009, 5. marraskuuta, saatavilla <http://www.w3.org/Protocols/>
- Http over tls*. (2000). IETF. Lainattu 2009, 16. marraskuuta, saatavilla <http://www.ietf.org/rfc/rfc2818.txt>
- Hypertext transfer protocol*. (1999). IETF. Lainattu 2009, 5. marraskuuta, saatavilla <http://www.ietf.org/rfc/rfc2616.txt>

- Internet usage statistics.* (2009, 30. syyskuu). Miniwatts Marketing Group.
Lainattu 2010, 23. helmikuuta, saatavilla
http://www.w3schools.com/browsers/browsers_stats.asp
- Jacobs, I. (toim.). (2004, 21. maaliskuu). *Uris, addressability, and the use of http get and post.* W3C. Lainattu 2010, 23. helmikuuta, saatavilla
<http://www.w3.org/2001/tag/doc/whenToUseGet.html>
- Järvinen, P., & Järvinen, A. (2004). *Tutkimustyön metodeista.* Tampere, Finland: Opinpajan Kirja. (ISBN:952-99233-2-5)
- Karjalainen, I. (2010). *Päätelaite- ja paikkatiedon automatisoitu kerääminen verkkopalveluissa* [Diplomityö].
- Kaushik, A. (2010). *Web analytics 2.0; the art of online accountability & science of customer centricity* (1. p.). Wiley Publishing, Inc. (ISBN:9780470529393)
- Kiker, G. A., Bridges, T. S., Varghese, A., Seager, T. P., & I, L. (2005). *Application of multicriteria decision analysis in environmental decision making. integrated environmental assessment and management 1(2).* 3909 Halls Ferry Road, Vicksburg, Mississippi 39180: U.S. Army Engineer Research and Development Center, Environmental Laboratory. Saatavilla
http://dx.doi.org/10.1897/IEAM_2004a-015.1
- Korpela, J. (2003, 28. syyskuu). *Methods get and post in html forms - what's the difference?* Lainattu 2010, 8. toukokuuta, saatavilla
<http://www.cs.tut.fi/~jkorpela/forms/methods.html>
- Kozierok, C. (2005). *The tcp/ip guide; a comprehensive, illustrated internet protocols reference* (1. p.). No Starch Press. Saatavilla
<http://books.google.fi/books?id=Pm4RgYV2w4YC&lpg=PP1&ots=qvkWjHMxdz&dq=TCP&pg=PP1#v=onepage&q=&f=false> (ISBN:159327047X)
- Krebs, V. (2008). *Social network analysis, a brief introduction.* orgnet.com.
Lainattu 2010, 5. toukokuuta, saatavilla <http://www.orgnet.com/sna.html>
- Laine, A. (2004). *Hahmolait käytettävyyden parantajina* [LuK-tutkielma]. Jyväskylän yliopisto, tietotekniikan laitos. Saatavilla
<http://www.mit.jyu.fi/opetus/opinnayte/LuK/Hahmolait/>
- Log analyzers comparisons.* (2009). AWStats. Lainattu 2009, 5. marraskuuta, saatavilla http://awstats.sourceforge.net/docs/awstats_compare.html
- Log files.* (s.a.). The Apache Software Foundation. Lainattu 2010, 23. helmikuuta, saatavilla <http://httpd.apache.org/docs/2.2/logs.html>

- Market share.* (s.a.). MySQL. Lainattu 2010, 9. toukokuuta, saatavilla
<http://www.mysql.com/why-mysql/marketshare/>
- McCaffrey, J. D. (2009). *Using the multi-attribute global inference of quality (magiq) technique for software testing*. Washington, DC, USA: IEEE Computer Society. (ISBN:978-0-7695-3596-8)
- Middleware.* (2010). Django. Lainattu 2010, 25. tammikuuta, saatavilla
<http://docs.djangoproject.com/en/dev/topics/http/middleware/>
- Mobiilisivut saavat kävijämittauksen. (2010, 26. helmikuu). , s. 32. Lainattu 2010, 8. toukokuuta, saatavilla
<http://www.kauppalehti.fi/5/i/yritykset/lehdisto/hellink/tiedote.jsp?direct=true&selected=kaikki&industry=&oid=20100301/12674340834360>
- Ncsa httpd.* (1998). NCSA. Lainattu 2009, 24. marraskuuta, saatavilla
<http://hoohoo.ncsa.illinois.edu/>
- Netissä roikutaan jo liki tunti päivässä. (2010, 29. huhtikuu). Lainattu 2010, 8. toukokuuta, saatavilla http://www.yle.fi/uutiset/kotimaa/2010/04/netissa_roikutaan_jo_liki_tunti_paivassa_1642518.html
- October 2009 web server survey.* (2009). Netcraft. Lainattu 2009, 5. marraskuuta, saatavilla http://news.netcraft.com/archives/web_server_survey.html
- Opera passes iphone to lead mobile-browser market. (2009, kesäkuu). Lainattu 2009, 16. marraskuuta, saatavilla
<http://www.reuters.com/article/marketsNews/idUSL224684120090602>
- Opera's history.* (2009). Opera. Lainattu 2009, 16. marraskuuta, saatavilla
<http://www.opera.com/company/history/>
- Oram, A. (2001). *Harnessing the power of disruptive technologies* (1. p.; D. Kelly, toim.). O'Reilly Media. Lainattu 2009, 13. marraskuuta, saatavilla
<http://oreilly.com/catalog/9780596001100> (ISBN:0-596-001110-X)
- O'Reilly, T. (2005, 30. syyskuu). *What is web 2.0; design patterns and business models for the next generation of software*. Lainattu 2010, 7. maaliskuuta, saatavilla <http://oreilly.com/web2/archive/what-is-web-20.html>
- Paul, R. (2009, 29. tammikuu). *Firefox share up over 20% in europe, mostly at expense of ie*. Lainattu 2009, 13. marraskuuta, saatavilla
<http://arstechnica.com/open-source/news/2008/01/firefox-gobbles-up-more-internet-explorer-market-share.ars>

- Popescu, A. (toim.). (2009, 7. heinäkuu). *Geolocation api specification; w3c working draft 07 july 2009*. W3C. Lainattu 2010, 20. helmikuuta, saatavilla <http://www.w3.org/TR/geolocation-API/>
- Shafranovich, Y. (2005). *Common format and mime type for comma-separated values (csv) files*. The Internet Engineering Task Force. Lainattu 2010, 26. maaliskuuta, saatavilla <http://tools.ietf.org/html/rfc4180>
- Sharma, P. (2008, 28. marraskuu). *Core characteristics of web 2.0 services*. Lainattu 2010, 7. maaliskuuta, saatavilla <http://www.techpluto.com/web-20-services/>
- Siirtola, H. (2007). *Interactive visualization of multidimensional data* [Väitöskirja]. Tampereen yliopistopaino Oy. Saatavilla <http://acta.uta.fi/haekokoversio.php?id=10949> (ISBN: 978-951-44-6871-1)
- Silius, K., Miilumäki, T., Huhtamäki, J., Tebest, T., Meriläinen, J., & Seppo, P. (2009, joulukuu). Social media enhanced studying and learning in higher education. IEEE EDUCON Education Engineering 2010 - The Future of Global Learning Engineering Education; April 14-16 2009, Madrid, SPAIN.
- Silius, K., Miilumäki, T., Huhtamäki, J., Tebest, T., Meriläinen, J., & Seppo, P. (2010, tammikuu). Students' motivations for social media enhanced studying and learning. Knowledge Management & E-Learning: An International Journal (KM&EL); Mar 2010, Vol.2, No.1.
- Sql null values*. (2010). W3C. Lainattu 2010, 26. tammikuuta, saatavilla http://www.w3schools.com/SQL/sql_null_values.asp
- Stevens, S. S. (1946). *On the theory of scales of measurement*. Science. Lainattu 2010, 19. helmikuuta, saatavilla http://web.duke.edu/philosophy/bio/Papers/Stevens_Measurement.pdf
- A tcp/ip tutorial*. (1991). IETF. Lainattu 2009, 5. marraskuuta, saatavilla <http://tools.ietf.org/html/rfc1180/>
- Timeplot; web widget for plotting time series*. (s.a.). Simile project MIT. Lainattu 2010, 23. helmikuuta, saatavilla <http://www.simile-widgets.org/timeplot/>
- Top 9 mobile browsers from dec 08 to now 09*. (2009). StatCounter Global Stats. Lainattu 2009, 16. marraskuuta, saatavilla http://gs.statcounter.com/#mobile_browser-ww-monthly-200812-200911

- Tufte, E. (2001). *The visual display of quantitative information* (2. p.). Graphics Press. (ISBN:0961392142)
- Valtionvarainministeriö. (2007, marraskuu). Verkkopalvelujen laatukriteeristö. Lainattu 2009, 16. marraskuuta, saatavilla http://www.suomi.fi/suomifi/laatuaverkkoon/laatukriteeristo/uusi_kriteeristo/verkkopalvelujen_laaturiteeristo/Verkkopalvelujen_laaturiteeristo.pdf (ISBN: 978-951-804-755-4)
- Verta, O.-M. (2006). *Päätösanalyysi vuorovaikutteisissa suunnittelun tukena; tapaustarkastelussa koitteen säännöstelyn monitavoitteinen kehittäminen* [Diplomityö]. Saatavilla <http://www.water.tkk.fi/wr/tutkimus/thesis/Verta2006.pdf>
- Vixie, P. (2007, toukokuu). *Dns complexity*. acmqueue. Lainattu 2009, 16. marraskuuta, saatavilla <http://queue.acm.org/detail.cfm?id=1242499>
- Ware, C. (2004). *Information visualization; perception for design*. Morgan Kaufmann Publishers.
- Web browser html support*. (2009). Web Devout. Lainattu 2009, 16. marraskuuta, saatavilla <http://www.webdevout.net/browser-support>