Jouni Laveri

Käyttäjien profilointi verkkokaupoissa.
Viitekehysmalli profilointimenetelmien vertailuun

Tietojärjestelmätieteen
Pro Gradu tutkielma
2.11.2006

Jyväskylän yliopisto
Tietojenkäsittelytieteiden laitos
Jyväskylä

TIIVISTELMÄ

Laveri, Jouni
Tietojärjestelmätieteen Pro Gradu tutkielma / Jouni Laveri
Ohjaaja: Anne Honkaranta
Jyväskylä: Jyväskylän yliopisto, 2006.
88 s.
Käyttäjien profilointi verkkokaupoissa. Viitekehysmalli profilointimenetelmien vertailuun

Käyttäjien profilointi antaa verkkokaupoille monia mahdollisuuksia erottautua kilpailijoistaan. Se tarjoaa kauppapaikoille työkalut luoda henkilökohtaista ja personoitua asiakaspalvelua. Käyttäjien profilointi on ollut viime vuodet varsin tutkittu aihe. Ehkä juuri tästä johtuen on tutkimuskenttä ollut varsin sekava niin termeiltään kuin menetelmiltäänkin.

Tässä työssä määritellään, mitä on käyttäjien profilointi verkkokaupoissa, mitkä ovat sen ominaisuudet ja kuinka profilointimenetelmiä voidaan vertailla. Profilointimenetelmien kenttään tehdään kirjallisuuskatsaus. Kirjallisuuskatsauksen tuloksena todetaan, että profilointi on perusteltua, mutta että se vaatii toimiakseen luottamusta. Kirjallisuuskatsauksen perusteella jaetaan profilointimenetelmät neljään osaan. Näiden tulosten pohjalta luodaan viitekehys profilointimenetelmien ja –järjestelmien vertailua varten. Lopuksi työssä laadittua viitekehystä verrataan kahteen muuhun jo olemassa olevaan viitekehykseen. Tämän työn viitekehyksen todetaan olevan vertailluista ainoa, joka keskittyy vertailemaan profilointimenetelmiä ja -järjestelmiä yleisesti. Molemmat muut viitekehykset keskittyvät vain johonkin tiettyyn profiloinnin osa-alueeseen.

AVAINSANAT: käyttäjien profilointi, tiedonsuodatus, personointi, verkkokaupat, asiakkuudenhallinta, CRM

ABSTRACT

Laveri, Jouni
Thesis for the degree of Master of Science (economics) / Jouni Laveri
Supervisor: Anne Honkaranta
Jyväskylä: University of Jyväskylä, 2006.
88 pages
User profiling in online shops. Framework model for profiling methods comparison

User profiling provides online shops many possibilities to differentiate themselves from competitors. It provides shops tools for personalizing services for customers. For the past few years user profiling has been a thoroughly researched subject. Maybe because of this, the research field's terms and methods have been unambiguous.

This thesis defines what user profiling is, what are it's features and how it is possible to compare different user profiling methods. A literature review is carried out for examining the research field. User profiling is found out to be well grounded, but it needs user's trust in order to function properly. Profiling methods are divided into four main categories. Based on the results of the literary review a framework for comparison of profiling methods and systems is presented. Finally the created framework is compared with two already existing frameworks. While the other two existing frameworks are focused on only one sub-category of profiling, the proposed framework concentrates profiling methods and systems in general.

KEYWORDS: User profiling, information filtering, personalization, online shops, CRM

SISÄLLYSLUETTELO

KUVALUETTELO

JOHDANTO

Www-ympäristössä liiketoimintaa harjoittavilla yrityksillä on valtava määrä potentiaalisia asiakkaita. Asiakkaita ei kuitenkaan tavata kasvokkain. Näillä yrityksillä ei näin ollen ole varsinaista henkilökohtaista kontaktia asiakkaaseensa. Asiakaskunta pitäisi näistä ongelmista huolimatta pystyä tuntemaan. Kuinka voidaan myydä tuotteita ilman, että tiedetään keitä asiakkaat ovat?

Yksi ratkaisu näihin ongelmiin on tässä työssä esiteltävä käyttäjien profilointi. *Käyttäjien profiloinnissa* (engl. user profiling) on kaksi vaihetta. Ensin selvitetään ne käyttäjän ominaisuudet, jotka ovat kauppapaikalle olennaisia ja muodostetaan näiden tulosten pohjalta käyttäjästä profiili. Tämän jälkeen profiilia hyödynnetään suodattamaan kauppapaikalla olevaa tietoa vastaamaan käyttäjän ominaisuuksia.

Käyttäjien profilointi on unelmatilanteessa hyvä ratkaisu niin yrityksille kuin niiden asiakkaillekin. Kun asiakkaat pystytään yhdistämään heitä kiinnostaviin tuotteisiin, asiakkaalta säästyy aikaa ja vaivaa ja samaan aikaan myös yrityksen myyntiluvut kasvanevat.

1.1 Tausta

Tässä työssä keskitytään esittelemään käyttäjien profilointia, jonka käyttöympäristönä on verkkokauppa. Aiemmissa tutkimuksissa profilointimenetelmiä on tutkittu lähinnä verkkokauppojen, hakukoneiden sekä uutisryhmien käytön apuna (esimerkiksi Wong ja Butz 2000, sekä Eirinaki ja Vazirgiannis 2003). Vaikka työ keskittyy tarkastelemaan profilointia verkkokauppojen näkökulmasta, ovat esitellyt menetelmät usein päteviä ja jopa suositeltavia myös muihin ympäristöihin, kuten edellä mainittujen hakukoneiden ja uutisryhmien käyttöön tai esimerkiksi tehostamaan yrityksen lähiverkkojen tiedonjakoa. Koska yritykset eivät mielellään luovuta liiketoimintansa kannalta olennaisia tietoja ulkopuolisille, on verkkokauppojen tutkiminen haastavaa.

Tutkielmassa käytetään www-ympäristössä toimivan yrityksen www-sivustosta termiä *kauppapaikka*. Jos on tarpeen korostaa sitä, että kyseeseen tulee mikä tahansa www-sivusto, käytetään termiä *sivusto*.

Tässä tutkimuksessa käyttäjien profilointi käsitetään laajemmaksi kokonaisuudeksi kuin personointi ja muokattavuus. Termi *personointi* (engl. personalization) tarkoittaa tässä yhteydessä käyttäjän kokemuksen henkilökohtaiseksi muokkaamista (Peppers, Rogers ja Dorf 1999). Termi *muokattavuus* (engl. customization) tarkoittaa tässä yhteydessä käyttäjän suorittamaa esimerkiksi kauppapaikan käyttöliittymän muokkausta (Allen, Kania ja Yaeckel 2001). Muokattavuus on osa personointia (Imhoff, Loftis ja Geiger 2001). Personointia voi olla vaikea erottaa käyttäjien profiloinnista. Esimerkiksi Cöner (2003) määrittää personoinnin olevan yrityksen suorittamaan toimintaa, jossa pyritään saattamaan profiloiduille käyttäjille luokiteltua sisältöä.

Yleiskäsite profiloinnin alle kuuluvat myös käsitteet profilointimenetelmä, profiili sekä profilointijärjestelmä. *Profilointimenetelmä* on menetelmä jonka tuloksia ovat profiili ja profiilin hyödyntäminen. *Profiili* taas sisältää tietoa asiakkaasta, jonka avulla saadaan lisäarvoa itse asiakkaalle, kauppapaikalle tai molemmille. *Profilointijärjestelmä* on jonkun tietyn kauppapaikan tai sivuston jotain profilointimenetelmää käyttävä järjestelmä (esimerkiksi Amazon.comin profilointijärjestelmä).

Muita profilointiin liittyviä termejä ovat suosittelujärjestelmät, tiedonsuodatus sekä tiedon louhinta. *Suosittelujärjestelmät* (engl. recommender systems) ovat nimensä mukaisesti järjestelmiä, jotka suosittelevat käyttäjälle sisältöä, jonka uskovat kiinnostavan häntä (esimerkiksi MovieLens). *Tiedonsuodatus* (engl. information filtering) on tiedon suodattamista niin että jäljelle jää vain olennainen (Belkin ja Croft 1992). *Tiedon louhinta* (engl. data mining) termiä käytetään yleensä tiedonsuodatuksen yhteydessä tarkoittamaan työkalua, jolla pyritään saamaan hyödyllistä informaatiota suuresta määrästä käsittelemätöntä tietoa

(Thuraisingham 2000). Tiedonsuodatusta voidaan pitää yläterminä tiedonlouhinnalle.

1.2 Rakenne

Luvussa kaksi luodaan kirjallisuuskatsaus käyttäjien profilointiin www-ympäristössä. Luvussa kuvataan erinäisiä syitä profiloinnin käyttöön kauppapaikan ja käyttäjän näkökulmasta. Tämän jälkeen kuvataan profiloinnin hyödyntämiskeinoja. Lopuksi käsitellään myös muita profilointiin liittyviä aiheita ja ongelmia, kuten käyttäjän tunnistamisen ja luottamuksen luomisen vaikeuksia.

Luvussa kolme luodaan kirjallisuuskatsaus erilaisiin käyttäjien profilointimenetelmiin. Luvussa käydään läpi profilointimenetelmien perustyypit sekä esitellään lyhyesti niiden pääpiirteet ja yleiset ongelmat. Aivan lopuksi tutustutaan profilointiin olennaisesti liittyviin tiedonsuodattamisen menetelmiin.

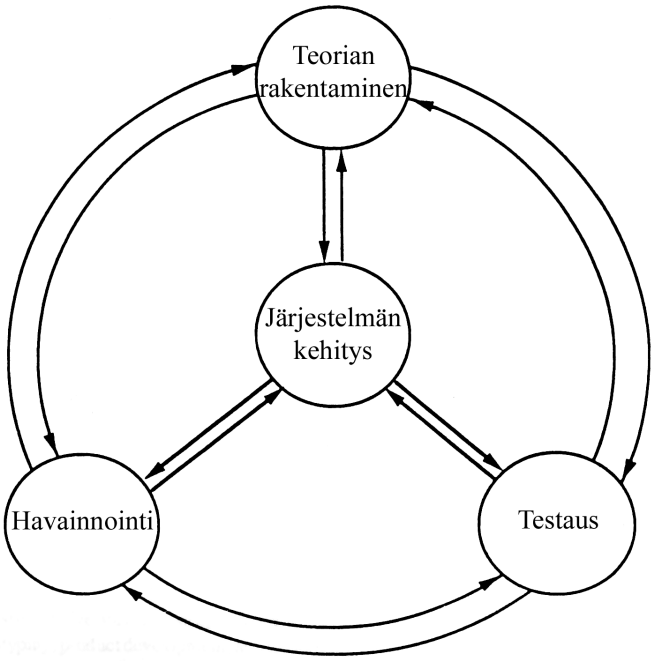
Luvussa neljä muodostetaan viitekehys profilointimenetelmien vertailua varten. Viitekehyksen pohjana käytetään lukujen kaksi ja kolme kirjallisuuskatsausten tuloksia. Luvussa kuvataan profilointimenetelmien yksittäisiä ominaisuuksia ja lopuksi liitetään nämä luotavaan Laverin viitekehykseen omiksi kokonaisuuksikseen.

Luvussa viisi verrataan Laverin viitekehystä kahteen muuhun profilointimenetelmien vertailukehykseen. Kaikkien kolmen viitekehyksen avulla tutkitaan MovieLens-nimistä elokuvien suosittelujärjestelmää. Luvun lopuksi vertaillaan näiden eri järjestelmien eroavaisuuksia saatujen tulosten perusteella.

1.3 Tutkimus

Työ on konstruktiivinen tutkimus. Tutkimusongelmana on: ”mitä on käyttäjien profilointi www-ympäristössä, mitkä ovat sen ominaisuudet ja kuinka profilointimenetelmiä voidaan vertailla?”

Tutkimusmenetelmänä käytetään osittain Nunamakerin, Chenin ja Purdinin (1990) esittämää mallia informaatioteknologian tutkimuksesta (kuva 1). Nuolet mallin osien välillä kuvaavat osien vaikutusta toisiinsa. Esimerkiksi saataessa havainnoinnissa lisää tietoa, voidaan sitä hyödyntää tarkentamaan teorian rakentamista. Lukujen kaksi ja kolme kirjallisuuskatsaus vastaa mallin kohtaa havainnointi. Luvun neljä viitekehyksen muodostaminen vastaa mallin kohtaa teorian rakentaminen. Viimeisenä luvun viisi viitekehyksen vertailu vastaa kohtaa testaus. Järjestelmän kehitys-vaihetta (engl. systems development) ei hyödynnetä tässä tutkimuksessa.



Kuva 1. Nunamakerin, Chenin ja Purdinin (1990) malli informaatioteknologian tutkimukseen

Havainnoinnissa (engl. observation) pyritään saamaan tutkittavasta alueesta yleiskuva, etenkin tilanteessa, jossa tietoa ei vielä suurissa määrin ole (Nunamaker ym. 1990, 95). *Teorian rakentamisessa* (engl. theory building) konstruoidaan esimerkiksi teoreettinen viitekehys (Nunamaker ym. 1990, 94). *Testaukses-*

sa (engl. experimentation) pyritään tutkimaan edellisten vaiheiden tuloksia ja mahdollisesti löytämään niistä ongelmia, joita voidaan myöhemmin korjata (Nunamaker ym 1990, 95).

2 PROFILOINTI

Internetillä oli maaliskuun lopussa 2006 yli miljardi käyttäjää maailmanlaajuisesti (Internet World Stats). Vaikka näistä käyttäjistä www:ssä olisi vain puolet, olisi sillä siitä huolimatta satoja miljoonia käyttäjiä. Mikään muu kauppa-alue ei pysty tavoittamaan potentiaalisesti näin paljon asiakkaita. Puhuttaessa tällaisista määristä mahdollisia asiakkaita nousee tärkeäksi menestystekijäksi erottautuminen muista yrityksistä. Jotta erottautuminen onnistuisi, tulisi asiakkaiden tuntea saavansa kaupasta jotain lisäarvoa verrattuna muihin jopa satoihin samankaltaisilta kauppoihin. Tähän tavoitteeseen päästäkseen olisi yrityksen tarpeellista pystyä hallitsemaan asiakkuuksiaan hyvin.

Kuinka sitten on mahdollista hallita tällaista määrää mahdollisia asiakkuuksia? Se ei luultavasti onnistu yhdessäkään kauppapaikassa manuaalisesti. Tätä hallintaa varten on kehitetty erilaisia *asiakkuudenhallinnanmenetelmiä* (CRM, Customer Relationship Management). Eräs tähän tarkoitukseen soveltuva menetelmä on käyttäjien profilointi. Profilointi voi tarkoittaa esimerkiksi käyttäjien tietojen keräämistä talteen kauppapaikassa säilytettävään profiiliin ja sopivan tuotteen tarjoamista tämän profiilin perusteella. Toisin sanoen käyttäjän erilaisista ominaisuuksista ja/tai mielenkiinnon kohteista muodostetaan kuva (eli profiili) hänestä kerättyjen tietojen pohjalta, käyttäen apuna mahdollisesti myös muiden käyttäjien tietoja. Tämän kuvan pohjalta pyritään saavuttamaan lisäarvoa joko käyttäjälle tai yritykselle. Tämä lisäarvo voi olla esimerkiksi edellä mainittuja asiakkaalle tarjottavia tuotesuosituksia tai yritykselle mahdollisuus kohdentaa mainontaansa oikealle ryhmälle.

Profilointimenetelmiä on kehitetty kahdesta eri näkökohdasta samanaikaisesti. Nämä ovat jo aiemmin mainittu asiakkuudenhallinnan profilointimenetelmät ja tämän lisäksi myös tiedonsuodatuksen profilointimenetelmät (Belkin ja Croft, 1992). Teknisesti tarkasteltuna profilointi on yleensä tiedonsuodatusta. Lähtölaukauksena profiloinnin kehittymiselle tiedonsuodatuksen puolella on voinut

toimia esimerkiksi sisäverkon tai vaikkapa hakukoneen tietojen järjestelyn tarve.

Näitä kahdesta eri lähtökohdasta kehitettyjä menetelmiä on vaikea erottaa toisistaan, koska menetelmien tavoitteet ovat yhtäläiset erilaisista lähtökohdistaan huolimatta. Samat menetelmät soveltuvat molempiin vaihtoehtoihin, joten tässäkään työssä ei ole näitä kahta lähtökohtaa eroteltu.

Tutkimukset, jotka käsittelevät käyttäjien profilointia, saattavat keskittyä myös profilointimenetelmien tiedonsuodatuspuoleen. Nämä tutkimukset käsittelevät usein joko hakukoneiden tai uutissivustojen profilointimenetelmiä, mutta käytetyn tekniikan homogeenisyyden ansiosta voidaan tuloksia hyödyntää käytännössä mille tahansa www-sivustolle.

Seuraavaksi tarkastellaan profiloinnin taloudellisia kysymyksiä. Tämän jälkeen esitellään lyhyesti joitakin tapoja hyödyntää käyttäjien profiileja niin yrityksen kuin käyttäjänkin näkökulmasta. Luku vastaa Nunamakerin, Chenin ja Purdinin (1990) esittämässä mallissa (katso 1.3) kohtaa havainnointi. Havainnoinnissa pyritään saamaan tutkittavasta alueesta yleiskuva, etenkin tilanteessa, jossa tietoa ei vielä suurissa määrin ole (Nunamaker ym. 1990).

2.1 Taloudellinen pohja profiloinnille

Www-ympäristö tarjoaa lähtökohtaisesti paremmat mahdollisuudet profiloinnille kuin perinteinen fyysinen ympäristö, koska kaikki asiakastapahtumat ovat jo valmiiksi digitaalisessa muodossa. Näin ollen asiakastiedon yhdistäminen profiiliksi on helpompaa kuin esimerkiksi paperimuodossa olevien asiakastietojen siirtäminen erikseen tietokoneelle. Joidenkin asiakastietojen siirto voi olla lähes mahdotonta. Asiakkaan kanssa esimerkiksi suullisesti kommunikoitaessa on vaikeaa muuttaa myyjän tietoja asiakkaasta digitaaliseen muotoon. (Vesanen ja Raulas 2006)

Perinteisessä ympäristössä on myös huomattu yritysten useasti pitävän erilaisia profilointiin liittyviä toimia yksittäisinä prosesseina. Kun näitä toimintoja ei yhdistetä kokonaisuudeksi, ei niistä myöskään saada täyttä hyötyä irti. (Vesanen ja Raulas 2006)

Käyttäjien profilointi www-ympäristössä tarjoaa yrityksille kaksi merkittävää etua. Ensinnäkin se antaa yrityksille mahdollisuuden tarjota asiakkailleen ajankohtaista ja tarkkaa tietoa tuotteistaan ja palveluistaan, näin ollen kasvattaen yrityksen myyntiä (Postma ja Brokke 2002). Tämän lisäksi profiloinnin on havaittu lisäävän asiakkaiden yritystä kohtaan tuntemaa uskollisuutta (Srinivasan, Anderson ja Ponnavolu 2002).

Yritykset voivat yrittää erottautua muista myös mahdollisimman henkilökohtaisen profiloinnin avulla, esimerkiksi tervehtimällä käyttäjää hänen nimellään (Wu, Tremaine, Instone ja Turoff 2002). On osoitettu, että yritysten voitot voivat nousta huomattavasti jos ne yrittäisivät nostaa osuuttaan yksittäisen asiakkaan suorittamista ostoksista sen sijaan, että nostaisivat asiakkaiden määrää (Danna ja Gandy 2002).

Profilointijärjestelmien tuottamien suositusten on todettu vaikuttavan asiakkaiden päätöksiin. Senecalin ja Nantelin (2004) tutkimuksen mukaan asiakkaat, joille oli suositeltu tuotteita, valitsivat suositellun tuotteen kaksi kertaa useammin kuin ryhmä, jolle suosituksia ei tarjottu. Vaikka asiakkaat pitivätkin profilointijärjestelmiä vähemmän asiantuntevina kuin ihmisasiantuntijoita tai toisia kuluttajia, vaikuttivat ne silti näistä kolmesta merkittävimmin asiakkaiden päätöksiin. (Senecal ja Nantel 2004)

Verkkokauppojen asiakkaat ovat tutkimusten mukaan heterogeenisiä eivätkä näin ollen eroa perinteisen fyysisen ympäristön asiakkaiden monimuotoisuudesta (Swinyard ja Smith 2003). Ilman jonkinlaista seurantaa ja jaottelua on vaikeaa, ellei mahdotonta tietää minkälaisia asiakkaita kauppapaikassa käy ja mil-

laisia tuotteita he haluavat ostaa. Profilointi tarjoaakin hyvät työkalut näiden keskenään heterogeenisten asiakkaiden toisistaan erottamiseen.

2.1.1 Asiakkaan vaihtokustannukset uskollisuustekijänä

Profiloinnin perimmäinen pyrkimys on saada asiakkaista uskollisempia. Asiakkaan uskollisuutta voidaan kasvattaa nostamalla hänen vaihtokustannuksiaan. *Vaihtokustannukset* (engl. switching costs) eivät välttämättä ole suoraan rahallisia kustannuksia, vaan esimerkiksi asiakkaan ajan ja vaivan kuluttamista. Näitä vaihtokustannuksia on kolmea eri tyyppiä: suorat vaihtokustannukset, mahdollisuuskustannukset sekä upotetut kustannukset. (Riemer ja Totz 2001).

Suorat vaihtokustannukset (engl. direct switching costs) ovat kustannuksia, jotka syntyvät kun asiakas vaihtaa uuteen kauppaan ja muodostaa tämän kanssa uuden sopimussuhteen (Riemer, Totz 2001). Näiden kustannusten osuus kasvaa kun kauppatapahtumien suuruus ja kesto kasvaa esimerkiksi yritykseltä yritykselle–kaupassa. Tämä johtuu siitä, että näissä tilanteissa kulutetaan enemmän aikaa oikean kumppanin löytämiseen ja suhteen muodostamiseen. Profilointi vähentää tuotteiden ja palveluiden välistä vertailtavuutta sekä lähentää asiakkaan ja kaupan suhdetta. Tällä tavalla lisätään asiakkaan suoria vaihtokustannuksia (Riemer ja Totz 2001).

Mahdollisuuskustannukset (engl. opportunity costs) tarkoittavat kustannuksia, jotka asiakkaalle syntyvät kun hän vaihtaa uuteen kauppaan ja menettää nykyisessä kaupassa olevat etunsa. Profilointi kasvattaa asiakkaan mahdollisuuskustannuksia kasvattamalla näitä etuja. (Riemer ja Totz 2001)

Upotetut kustannukset (engl. sunk costs) ovat monessa mielessä lähellä mahdollisuuskustannuksia. Upotetut kustannukset ovat asiakkaan etujen saamiseen laittamat ajalliset tai rahalliset resurssit. Näiden kustannusten merkitys voi olla monessa suhteessa ennemminkin psykologinen kuin varsinaisesti rahallinen. Tällainen kustannus voisi olla esimerkiksi henkilökohtaisen profiilin kehittämi-

seen kulutettu aika. Profilointi antaa mahdollisuuksia näiden kustannusten syntymiseen parantaen näin asiakkaiden uskollisuutta. (Riemer ja Totz 2001)

2.1.2 Pienten ja keskisuurten yritysten näkökulma

Sveitsissä tehdyssä tutkimuksessa (Schubert ja Leimstoll 2004) tarkasteltiin pienten ja keskisuurien yritysten suhtautumista elektronisessa liiketoiminnassa tapahtuvaan profilointiin. Asiakkaan henkilökohtaisen huomioimisen todettiin olevan yrityksen menestymiselle elintärkeää (Schubert ja Leimstoll 2004). Tämä tulos oli pätevä myös erillään verkkoliiketoiminnasta. Www-ympäristössä tapahtuva profilointi antaa kuitenkin monia työkaluja tähän henkilökohtaiseen suhteeseen.

Tutkimus paljasti, että suurin osa pienistä ja keskisuurista yrityksistä ei omista omia www-palvelimia vaan vuokraa niitä ulkopuoliselta palveluntarjoajalta. Tästä johtuen tutkimuksessa pääteltiin, että ainut järkevä tapa profilointimenetelmien tehokkaaseen käyttöön olisi saada nämä palveluntarjoajat mukaan kehittämään profilointia. (Schubert ja Leimstoll 2004)

Edellä mainitussa tutkimuksessa olleissa yrityksissä on käytössään keskenään huomattavastikin eroavia järjestelmiä. Tämä heterogeenisyys tuottaa ongelmia profilointijärjestelmien käyttöönotossa, koska järjestelmät vaativat erillistä muokkausta jokaista järjestelmäalustaa varten. (Schubert ja Leimstoll 2004)

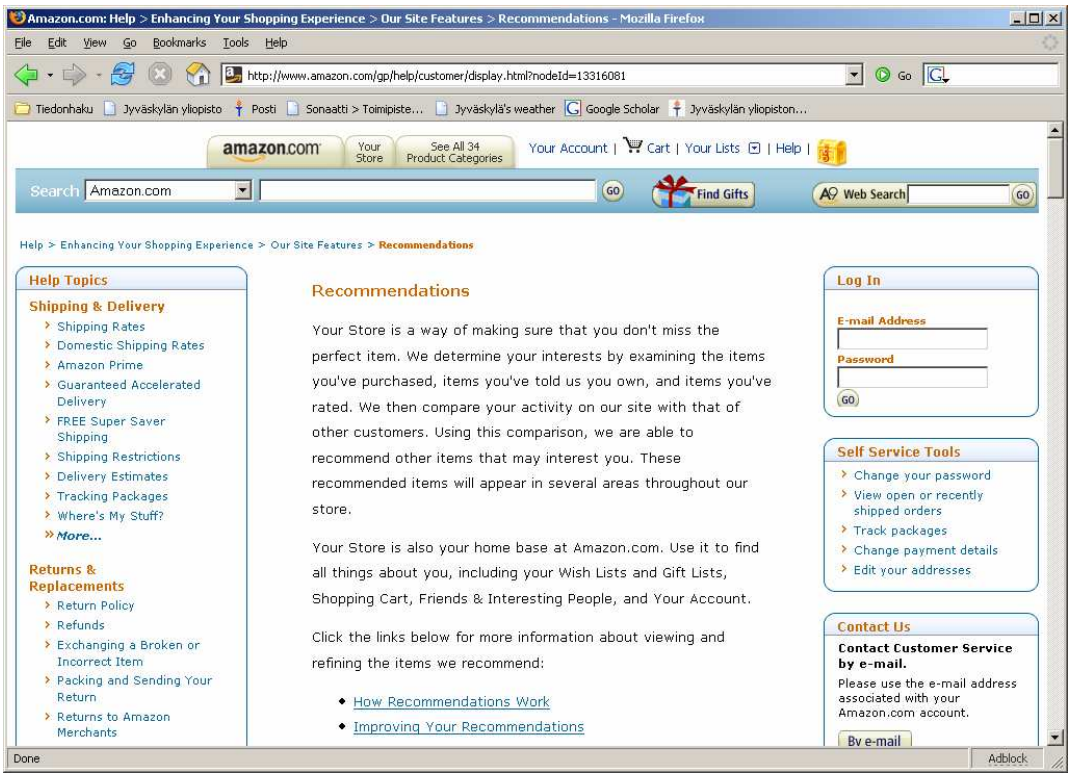
Näistä edellä mainituista ongelmista huolimatta tutkituista huomattava osa yrityksistä oli suunnittelemassa mittavia sijoituksia profilointiin ja profilointijärjestelmiin muutaman seuraavan vuoden aikana (Schubert ja Leimstoll 2004).

2.1.3 Profiloinnin läpinäkyvyys asiakkaille

Läpinäkyvyydellä tarkoitetaan tässä sitä, miten helposti asiakkaat voivat saada selville yrityksen erilaisia toimintoja. Nämä toiminnot voivat olla mitä tahansa

yrityksen liiketoimintaan liittyviä prosesseja. Yleisesti asiakkaita kiinnostavia toimintoja ovat kuitenkin ne, jotka jollain tavalla koskettavat häntä itseään. (Awad ja Krishnan 2006)

Kuvassa 2 näkyy Amazon.com kauppapaikan kuvaus sen asiakkailleen tarjoamasta profilointipalvelusta. Kuvassa näkyy yksi esimerkki siitä miten profiloinnista voi tehdä läpinäkyvää asiakkaalle. Amazon.com kertoo, mitä tuotesuositukset (kuvassa recommendations) ovat, miksi niitä kannattaa käyttää ja miten ne pääperiaatteiltaan toimivat. Amazon.com on yksi tunnetuimmista profilointimenetelmiä hyödyntävistä yrityksistä.



Kuva 2. Profilointipalvelun kuvaus (Amazon.com)

Miten läpinäkyvää asiakkaiden profiloinnin sitten pitäisi olla? Awadin ja Krishnanin (2006) suorittamassa tutkimuksessa havaittiin, että asiakkaat, jotka

haluavat suurempaa läpinäkyvyyttä profilointiin ovat samoja, jotka yleisesti ovat haluttomampia osallistumaan profiilin muodostukseen. Tutkimuksen mukaan nämä läpinäkyvyyttä haluavat asiakkaat ovat kuitenkin vain pieni osa kaikista asiakkaista.

Awadin ja Krishnanin (2006) tutkimuksen tulosten valossa voi päätellä, ettei yrityksen käyttämän profilointiprosessien tarvitse olla asiakkaille läpinäkyviä. Tämä siksi, että vaikka prosessit olisivat läpinäkyviä, profilointiin osallistuu parhaimmillaankin vain marginaalisesti suurempi joukko asiakkaita. Tämän lisäksi saatetaan menettää kilpailuetuja paljastamalla yrityksen menetelmiä julkisesti, puhumattakaan mahdollisesti aiheutuvista tietoturvariskeistä. Monet kauppapaikat eivät annakaan julkisuuteen juuri mitään tietoja profiloinninmenetelmistään (esimerkiksi Amazon.com), vedoten esimerkiksi juuri kilpailuetuun (Middleton 2003, 20).

2.2 Tapoja profiilin hyödyntämiseen

Miten profiilia voidaan sitten hyödyntää käytännössä? Tätä voidaan tutkia esimerkiksi asiakkaan ja tiedonkeruun näkökulmasta (Wu ym. 2002). Asiakkaan näkökulmasta tarkasteltuna profiloinnin hyödyntäminen saattaa olla esimerkiksi käyttöliittymän tai toimintojen muokkausta asiakkaan haluamalla tavalla. Tiedonkeruun näkökulmasta tarkasteltuna hyödyntämistavat saattavat olla paljon näkymättömämpiä, kuten hakutulosten järjestäminen asiakkaan profiilia vastaavalla tavalla.

Toinen tapa jakaa profiilin hyödyntämistapoja on jakaa ne viestintään, kauppapaikkaan tai tuotteisiin ja palveluihin liittyviin hyödyntämistapoihin (Riemer ja Totz 2001). Tämä jako on suoritettu tutkimalla järjestelmän ominaisuuksia lähinnä asiakkaan näkökulmasta. Tästä huolimatta se käy aivan hyvin myös yleisesti profilointijärjestelmien tarkasteluun.

Seuraavassa esitellään Wuta ym. (2002) mukaillen tehty jaottelu erilaisiin hyödyntämistapoihin. Jaottelua täydennetään Riemerin ja Totzin (2001) esittämillä ominaisuuksilla. Koska kauppapaikan näkökulmaan sisältyy myös tiedonkeruun näkökulma, on Wun ym. mallissa esiintyvä tiedonkeruun näkökulma korvattu kauppapaikan näkökulmalla. Wun ym. (2002) lähestymistapa aiheeseen on personointiläheinen, mutta niin on myös suurin osa profiilin hyödyntämistavoista. Tässä esiteltävät hyödyntämistavat eivät ole myöskään ainoita mahdollisia, mutta antavat hyvän kuvan siitä, minkälaisin tavoin profiilia voidaan hyödyntää. Ensin esitellään hyödyntämistavat kauppapaikan näkökulmasta ja sitten hyödyntämistavat asiakkaan näkökulmasta.

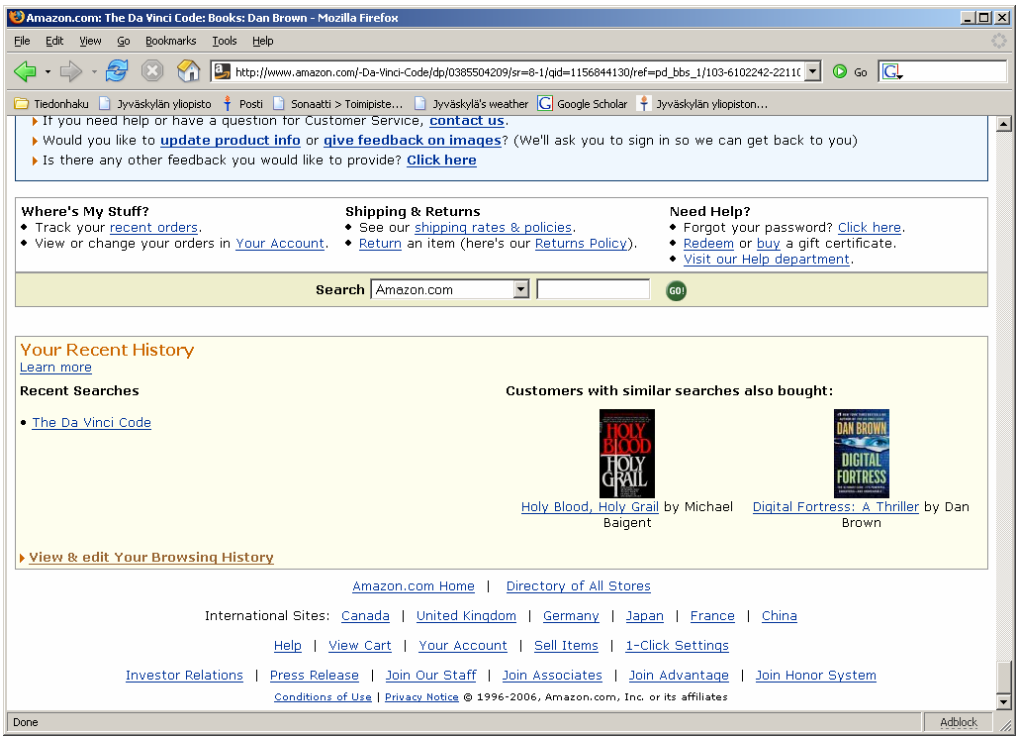
2.2.1 Hyödyntämistavat kauppapaikan näkökulmasta

Henkilökohtaiset työkalut (engl. personal tools) ovat asiakkaan itsensä luomia linkkejä kauppapaikassa oleviin, häntä kiinnostaviin tietolähteisiin. Esimerkiksi osakkeisiin liittyviä palveluja tarjoavan kauppapaikan sivustoilla asiakas voisi luoda linkkejä sivuille, joilla seurataan yhtäaikaisesti häntä kiinnostavien osakkeiden pörssikurssien kehitystä. (Wu ym. 2002)

Opportunistiset linkit (engl. opportunistic links) ovat linkkejä, jotka ilmestyvät näkyviin liittyen johonkin asiakkaan suorittamaan toimintoon. Asiakas voi esimerkiksi varata lentolipun, ja vahvistamisen yhteydessä hän saa näkyviin määränpäässä sijaitsevien hotellien ja nähtävyyksien mainoksia (Wu ym. 2002). Tällaisten opportunististen linkkien tehokkuutta voidaan lisätä liittämällä niihin muita profiloinnin keinoja.

Suosittelujärjestelmät ovat perinteisesti olleet profilointimenetelmistä tehokkaimpia asiakastiedon käsittelijöitä. Nimensä mukaisesti nämä järjestelmät suosittelevat asiakkaille tuotteita tai palveluita, perustuen joko asiakkaan tai samalla tavalla käyttäytyneiden asiakkaiden aikaisempiin ostoksiin (Wu ym. 2002). Nämä järjestelmät ovat varsinaisia profilointijärjestelmiä. Kuvassa 3 näkyy

Amazon.comin järjestelmä suosittamassa asiakkaalleen muita kirjoja, joita muut samaa tuotetta etsineet asiakkaat ovat ostaneet (”customers with similar searches also bought”). Suositukset muodostetaan vertailemalla asiakkaiden tietoja keskenään.



Kuva 3. Suosittelujärjestelmä. (Amazon.com)

Henkilökohtainen hinnoittelu (engl. individualized pricing) tarkoittaa nimensä mukaisesti asiakkaan profiilin mukaan suoritettua hinnoittelua (Riemer ja Totz 2001). Esimerkiksi lentoyhtiöllä voi olla ohjelma, jossa jonkun tietyn lentokilometrimäärän jälkeen saa alennuksia tulevaisuudessa maksettavista lennoista. On hyvä huomata, että henkilökohtainen hinnoittelu ei sinänsä eroa normaalisti käytössä olevista myynnintavoista, mutta yksinkertaistuu huomattavasti kun kaikki voidaan tehdä automaattisesti. Tällaisessa hinnoittelupolitiikassa täytyy kuitenkin olla varovainen, etteivät ”huonommat” asiakkaat tunne itseään eriar-

voiseksi ”parempien” asiakkaiden kanssa. Tosin tällaista eriarvoisuuttakin voidaan mahdollisesti käyttää hyväksi, jos asiakkaat saadaan tavoittelemaan itsellensä parempaa asemaa asiakashierarkiassa esimerkiksi kertomalla kanta-asiakaseduista.

2.2.2 Hyödyntämistavat asiakkaan näkökulmasta

Seuraavassa tarkastellaan profiilin hyödyntämistapoja asiakkaan näkökulmasta. Asiakkaan näkökulmasta profiloinnin hyödyntämistavat ovat usein personointia. Nämä hyödyntämistavat ovat asiakkaalle suoraan näkyviä ja mahdollisesti myös heidän suoraan hallittavissaan. Seuraavaksi esiteltävät hyödyntämistavat saattavat olla osittain päällekkäisiä, eivätkä yleensä sulje toisiansa pois. Nämä tavat saattavat olla osittain päällekkäisiä myös edellisessä kohdassa (2.2.1) esiteltyjen hyödyntämistapojen kanssa.

Profiloinnin hallinta (engl control personalization) viittaa yleisesti siihen, kuinka paljon asiakkaalla on valtaa henkilökohtaiseen personointiinsa. Asiakkaalle voidaan antaa esimerkiksi valta pysäyttää hänelle tulevat suositukset tai mahdollisesti välttää kokonaan profilointiprosessiin osallistuminen. Asiakkaan kokeman hallinnan tunteen on todettu lisäävän asiakkaiden tyytyväisyyttä huomattavasti. (Wu ym. 2002)

Sisällön personointi (engl. content personalization) viittaa kauppapaikan sisällön muuttamiseen profiilin mukaan. Sisällön personointi voi olla esimerkiksi Amazon.comissa tapahtuvaa tuotteiden suosittelemista sen mukaan, mitä aikaisemmin kauppapaikalla asioineet asiakkaat ovat ostaneet, kuten kuvassa 3 havainnollistettiin. (Wu ym. 2002, Riemer ja Totz 2001)

Linkkien personoinnissa (engl. link personalization) pyritään karsimaan tai parantamaan navigointi- ja tuotelinkkejä asiakaskohtaisesti (Wu ym. 2002). Vaikka menetelmä toimii myös ilman profilointia, voidaan sitä tehostaa huomattavasti ajan tasalla olevan profiilin avulla.

Asiakaskohtainen ulkoasun personointi (engl. customized screen design personalization) viittaa asiakkaan mahdollisuuteen muuttaa kauppapaikan ulkonäköä haluamaansa suuntaan. Tämä voi olla esimerkiksi värien tai asettelun muuttamista sivustolla tai jopa oman kalenterin luomista kauppapaikkaan. (Wu ym. 2002, Riemer ja Totz 2001)

Ihmismäinen käyttöliittymä (engl. anthropomorphic personalization) viittaa yksinkertaisesti siihen, että käyttöliittymä yrittää luoda ihmismäistä vuorovaikutusta asiakkaan ja kauppapaikan välille. Tämä voi tapahtua esimerkiksi tervehtimällä asiakasta nimeltä hänen saapuessaan kauppapaikkaan. (Wu ym. 2002)

Markkinointiviestien (esim. sähköpostit ja ponnahdusikkunat) lähettämistä asiakkaalle kannattaa harkita tarkoin (Riemer ja Totz 2001). Jos asiakas kokee saavansa roskapostia, ja etenkin jos hän ei pysty vaikuttamaan sen vastaanottamiseen, saattaa tuloksena olla negatiivinen mielikuva yritystä kohtaan. Suomessa suoramarkkinointi on myös lailla säädeltyä (Sähköisen viestinnän tietosuojalaki 516/2004).

2.3 Käyttäjän tunnistus

Käyttäjän tunnistaminen on olennainen osa profiilin toimintaa. Ilman luotettavaa tapaa käyttäjän tunnistamiseen ei ole mitenkään mahdollista luoda tai ylläpitää profiilia. Profiilin tunnisteelle on perinteisesti ollut kaksi eri tallennuspaikkaa. Ensimmäinen on käyttäjän koneella evästeen muodossa säilytettävä tunniste (2.3.1). Toinen on sisäänkirjautumisen avulla suoritettava tunnistus (Cingil 2002) (2.3.2). Näillä molemmilla tavoilla on kuitenkin ongelmansa. Näitä ongelmia käsitellään alla tarkemmin. Cingil (2002) on ehdottanut näihin mainittuihin ongelmiin ratkaisuksi kolmatta tunnistamisen keinoa nimeltään ”luotettu taho” (2.3.3).

2.3.1 Eväste

Yleisesti käytössä oleva käyttäjän tunnisteen säilytyspaikka on eväste. *Evästeet* (engl. cookie) ovat pieniä käyttäjän selaimeen tallennettavia tiedostoja. Asiakkaan tullessa ensi kertaa kauppapaikkaan tallennetaan hänen koneelleen eväste. Tämä eväste ei yleensä sisällä muuta kuin tunnuksen, jonka avulla osataan yhdistää kauppapaikan palvelimella olevat tiedot oikeaan henkilöön (Wu ym. 2002).

Eväste on kuitenkin osoittautumassa huonoksi säilytyspaikaksi lähinnä kahdesta pääongelmasta johtuen. Ensimmäinen ”ongelma” ovat yleistyvät haittaohjelmienpoistajat. *Haittaohjelmat* (engl. adware/spyware) ovat ohjelmia, jotka esimerkiksi pakottavat käyttäjän katselemaan mainoksia tai vakoilevat tämän salasanoja vaikkapa nettipankkiin. Käyttäjät saattavat poistaa evästeitä myös omatoimisesti haittaohjelmienpoistajien opettamana tai tietoturvaongelmien pelottamana. (Cingil 2002)

Toinen suuri ongelma on saman käyttäjän päätelaitteiden monimuotoisuus (Cingil 2002). Käyttäjällä saattaa olla useita eri laitteita, joilla hän kirjautuu esimerkiksi www-sivulle. Näihin voivat kuulua esimerkiksi PDA, kännykkä, pöytäkone ja kannettava tietokone. Kerättäessä profiilia perinteisellä tavalla vain yhdelle näistä laitteista, ei se ole mitenkään saatavilla muita laitteita käytettäessä.

2.3.2 Sisäänkirjautuminen

Toinen tapa käyttäjän tunnistamiseen on tunnisteen sijaitseminen kauppapaikan palvelimella ja käyttäjän tunnistaminen sisäänkirjautumisella. Tämä on evästeeseen verrattuna huomattavasti suositeltavampi tapa, koska tunnistaminen on varmempaa. (Cingil 2002)

Vaikka rekisteröityminen onkin käytössä monessa kauppapaikassa, monesti tätä mahdollisuutta ei kuitenkaan hyödynnetä profiilin muodostamisessa. Ongelmana menetelmässä saattaa olla myös käyttäjien haluttomuus luovuttaa tarkkoja tietoja itsestään, jolloin tarkan profiilin muodostamisesta tulee hankalaa, ellei jopa mahdotonta.

2.3.3 Luotettu taho

Cingil (2002) ehdottaa ratkaisuksi profiilien jaettavuusongelmaan ”*luotettua tahoa*” (engl. trusted authority). Tämän tahon hallinnassa sijaitsisivat käyttäjien profiilit. Nämä profiilit luovutettaisiin ainoastaan erikseen varmistetuille tahoille. ”Luotettu taho”-menetelmä pohjautuu seuraaville standardeille ja suosituksille: RDF (Herman, Swick ja Brickley 2006), XML (Extensible Markup Language (XML) 1.0 (Fourth Edition) 2006), XML-QL (Deutsch, Fernandez, Florescu, Levy ja Suciu 1998), sekä P3P (Platform for Privacy Preferences (P3P) Project 2006). Seuraavaksi esitellään tämä ratkaisu tarkemmin Cingiliä (2002) mukaillen.

P3P (Platform for Privacy Preferences) on W3C:n suositus, jonka tarkoituksena on helpottaa sekä selkeyttää käyttäjien ja kauppapaikkojen välistä tietosuojasopimuksen muodostamista. Se sisältää perusmuodossaan monivalintakysymyksiä, jotka kattavat verkkotietoturvallisuuden kaikki osat. Käyttäjä vastaa näihin kysymyksiin ja tämän perusteella luodaan ohjelmallisesti luettavissa oleva tietosuojaprofiili käyttäjälle. Kun käyttäjä tulee uuteen kauppapaikkaan, joka tukee P3P:ta, verrataan käyttäjän tietosuojaprofiilia kauppapaikan omaan *tietosuojaselosteeseen* (engl. privacy policy) tarkistaen onko kyseinen seloste tarpeeksi tiukka käyttäjän antamille vaatimuksille. Näin käyttäjän ei tarvitse joka kerta hyväksyä tietosuojasopimuksia uudestaan. (Cingil 2002, sekä Platform for Privacy Preferences (P3P) Project 2006)

”Luotettu taho”-menetelmässä käytetään selaustietoa samalla tavalla kuin aiemmin esitellyssä käytönseurantapohjaisessa profiloinnissa. Itse asiassa nämä kaksi menetelmää ovat tältä osin samoja, ainoastaan hyödyntämistapa tekniikoiden osalta eroaa. ”Luotettu taho”-menetelmässä tallennetaan käyttäjän selainkäyttäytyminen XML-tiedostoon samalla kun hän käyttää www:iä. Tästä syntyvä tiedosto on kuitenkin käsittelemätöntä tietoa, joka ei ole vielä sellaisenaan hyödyllistä. XML-tiedoston dokumenttityyppimäärityksen (DTD tai skeema) tulisi olla yhteinen kaikille, jotta sitä pystytään hyödyntämään yhteisten profiilien muodostamiseen eri kauppapaikoissa. (Cingil 2002)

Edellisessä vaiheessa saatuun XML-tiedostoon sovelletaan erilaisia tiedonsuodatusmenetelmiä, jotta saataisiin profiilin kannalta hyödyllistä informaatiota. Cingilin menetelmässä tämä informaatio poimitaan XML-tiedostosta XML-QL – kyselykieltä hyödyntäen. Tiedonsuodatusmenetelmistä kerrotaan tarkemmin luvussa kolme. Kun XML-tiedostosta on suodatettu profiilin kannalta olennainen tieto, tallennetaan tämä metatieto RDF-tiedostoon. (Cingil 2002)

”Luotettu taho” itsessään on erikoistunut *verkkopalvelu* (engl. web service), joka hoitaa käyttäjien profiilien, sekä tietosuojasopimusasetusten hallinnoinnin. Käyttäjät voivat lisätä profiileihin halutessaan myös tietoa, jota ei normaalisti saa selainkäyttäytymistä seuraamalla, esimerkiksi ammatin, perhetaustan sekä tulojen määrän. (Cingil 2002)

Käynnistäessään selaimen käyttäjän tulisi rekisteröityä käyttämäänsä luotettuun tahoon ja ladata profiilinsa esimerkiksi selaimen välimuistiin. Tietosuojaa parantaakseen voi luotettu taso tallentaa profiilit *salatussa* (engl. encrypted) muodossa. Optimaalisessa tapauksessa kaikki tämä olisi mahdollisimman automatisoitua, jotta vältyttäisiin käyttäjän kuormittamiselta tarpeettomasti. (Cingil 2002)

Standardeja ja suosituksia (XML, RDF, sekä P3P) käyttämällä parannetaan yhteensopivuutta eri kauppapaikkojen, selainten sekä päätelaitteiden välillä.

”Luotettu taho” menetelmä ratkaisisi suuren osan ongelmista, joita liittyy profiilin siirtämiseen eri päätelaitteiden välillä. (Cingil 2002)

2.4 Luottamus

Profiilin muodostukselle on kriittisen tärkeää, että kauppapaikka pystyy keräämään ja käsittelemään tietoa asiakkaista(Chellappa ja Sin 2005). Asiakkaiden täytyy myös luovuttaa näitä tietoja itsestänsä (Chellappa ja Sin 2005). Eräs suurimmista profiilien menestymisen ongelmista onkin käyttäjien pelko tietosuojansa puolesta, jolloin he eivät suostu luovuttamaan tietoja itsestänsä (Cingil 2002). Tämä ei sinänsä ole aiheeton pelko. Nykyään www:iä käyttävä joutuu sivuille rekisteröityessään miettimään tarkkaan, mitä tietoja luovuttaa. Esimerkiksi harvalle kauppapaikalle tai muullekaan sivustolle kannattaa antaa henkilökohtaista sähköpostiosoitettaan.

Satunnaisten käyttäjien toimintaa on lähes mahdoton seurata, koska he eivät ole valmiita rekisteröitymään kauppapaikkoihin. Käyttäjien pitäisikin kokea heidän saamansa palkkion olevan tarpeeksi suuri antaakseen luvan käyttäytymisensä seuraamiseen. Ilman luottamusta monet kokevat tämän seuraamisen jopa yksityisyyden loukkaamisena. (Middleton, 2003, 24)

Yrityksen tulee olla todella tarkkana, minkälaisen kuvan se antaa profilointiin tarvittavien tietojen keräämisestä. Profiloinnin ei pidä vaikuttaa vakoilulta. Käyttäjällä olisi hyvä olla mahdollisuus kieltää tietojen kerääminen niin halutessaan. Jo pelkästään tämä mahdollisuus kieltämiseen saattaa olla riittävä luottamuksen luoja, jotta käyttäjä antaa oikeaa tietoa itsestään profiilin kerättäväksi, tai mahdollisesti käyttää kauppapaikkaa ylipäätään. Asiakkaat, jotka kokevat, että heistä kerätään tietoja ilman lupaa ja näitä kerättyjä tietoja käytetään väärin, saattavat jopa yhdistää voimansa näitä yrityksiä vastaan. Profiloinnin yleistyessä yksityisyydestään huolestuneet asiakkaat ovatkin perustaneet muiden muassa CASPIAN (Consumers Against Supermarket Privacy Invasion and Numbering) nimisen painostusjoukon (Risch ja Schubert 2005).

Tähän luottamuspulaan saattaa antaa hiukan apua edellä esitelty (2.3.3) ”luotettu taho” menetelmä, mutta ongelmaa se ei ratkaise. Menetelmän apu piilee siinä, että käyttäjä näennäisesti rekisteröityy ainoastaan profiilinsa säilyttävälle sivustolle tai kauppapaikkaan. Tämän lisäksi käyttäjän luottamusta lisää se, että kauppapaikat, joihin “luotettu taho” menetelmää voisi soveltaa, tulisi olla tarkastettuina kyseisen tahon toimesta ennen luotettuun listaan mukaan hyväksymistä. Tämä kaikki lisää osaltaan käyttäjien luottamusta yleisesti www:iä, ja etenkin profiileja kohtaan, mutta ei täysin poista ongelmaa luottamuspulasta.

Profilointia on pidetty yleisesti luottamusta kasvattavana tekijänä. Jotkut tutkimukset ovat kuitenkin päätyneet jopa päinvastaiseen tulokseen, osoittaen että alkuun profilointi saattaa jopa heikentää luottamusta kauppapaikkaan (Serino ja Furner 2005). Alkuvaiheessa onkin hyvä kertoa avoimesti, että miksi kauppapaikka esimerkiksi suosittelee joitain tuotteita kyseiselle asiakkaalle (Adomavicius ja Tuzhilin 2005).

Oikeastaan ainut varmasti toimiva ratkaisu luottamuspulaan on yksinkertaisesti luoda kauppapaikalle luotettava kuva. Aikaisemmat tutkimukset ehdottavat tähän kahta eri tapaa. Ensimmäiseksi kauppapaikan tulisi muokata brändiänsä luotettavammaksi (Ward ja Lee 2000). Toiseksi voidaan tehdä yhteistyötä luotettavien kolmansien osapuolten kanssa ja tätä kautta nostaa kauppapaikan luotettavuutta (Friedman, Kahn ja Howe 2000). Kun käyttäjien mieliin on saatu muodostettua luotettava kuva kauppapaikasta, ei heidän tarvitsekaan enää luottaa yleisesti www:iin, kunhan he vain luottavat kyseiseen kauppapaikkaan. Jos näin on, antavat käyttäjät paremmin totuutta vastaavia tietoja itsestänsä, ja profiilinmuodostaminen on mahdollista ja myös huomattavasti tarkempaa.

Tässä luvussa tutkittiin, mitä profilointi on ja mitkä syyt tukevat sen käyttöä. Syitä löytyi luottamuksen luomisesta vaihtokustannuksiin. Luvussa esiteltiin myös useita profiloinnin hyödyntämistapoja niin asiakkaan kuin kauppapaikankin näkökulmasta. Tämän jälkeen käsiteltiin käyttäjien tunnistamisen ja pro-

filoinnin tarvitseman luottamuksen luomisen ongelmia. Käyttäjien tunnistamiseen löydettiin kolme erilaista tapaa: evästeet, sisäänkirjautuminen ja kolmas osapuoli. Luottamuksen luonnissa ainoan todella toimivan ratkaisun todettiin olevan luotettavan brändin luominen kauppapaikalle. Seuraavassa luvussa tarkastellaan lähemmin profilointimenetelmien perustyyppejä ja niiden ominaisuuksia sekä tehdään alustava jako erilaisten profilointimenetelmien kesken.

3 PROFILOINTIMENETELMÄT

Tässä luvussa luodaan kirjallisuuskatsaus olemassa oleviin profilointimenetelmiin, jotta saataisiin yleiskuva profilointimenetelmien ominaisuuksista. Luvussa esitellään alustava jako profilointimenetelmien kesken. Tämän luvun tuloksia käytetään pohjana luvussa neljä luotavalle viitekehykselle. Aivan kuten luku kaksi, vastaa tämäkin luku Nunamakerin, Chenin ja Purdinin (1990) esittämässä mallissa (katso 1.3) kohtaa havainnointi. Havainnoinnissa pyritään saamaan tutkittavasta alueesta yleiskuva, etenkin tilanteessa, jossa tietoa ei vielä suurissa määrin ole (Nunamaker ym. 1990, 95).

Profilointijärjestelmiä on lukuisia erilaisia (esimerkiksi Amazon.comin ja MovieLensin järjestelmät). Nämä järjestelmät ovat useimmiten kuitenkin vain muunnelmia muutamasta perusmenetelmästä. Erilaisia profilointimenetelmiä on tutkittu laajasti (esimerkiksi Middleton 2003, Riemer ja Totz 2001). Monet tutkimuksista ovat kuitenkin keskittyneet vain johonkin tiettyyn menetelmään, eikä ole niinkään kartoitettu kokonaisuutta. Pääkohteena on useissa tutkimuksissa ollut www:ssä tehtävä käyttäjien profilointi. Kohteina on ollut esimerkiksi sisällön suodattaminen verkkokauppojen, hakukoneiden tai uutisryhmien yhteydessä.

Seuraavaksi käydään läpi menetelmien perustyypit, sekä esitellään lyhyesti niiden peruspiirteet ja yleiset ongelmat. Tämän jälkeen käsitellään muita profilointiin liittyviä aiheita. Näitä ovat käyttäjän tunnistaminen sekä käyttäjien verkkokauppaa kohtaan kokema luottamus. Aivan lopuksi tutustutaan profilointiin olennaisesti liittyviin tiedonsuodattamisen menetelmiin.

3.1 Profilointimenetelmien perustyypit

Profilointimenetelmiä voidaan luokitella esimerkiksi sen tuotoksen eli profiilin mukaan. Profilointimenetelmän luoma profiili voi olla *dynaaminen* tai *staattinen* (Eirinaki ja Vazirgiannis 2003). Staattinen profiili on täysin tai lähes muuttuma-

ton profiili, johon ei luultavasti kosketa muodostamisen jälkeen lainkaan. Staattinen profiili voi olla esimerkiksi demografiseen tietoon pohjaava eli profiili voi koostua muun muassa seuraavista tiedoista: käyttäjän sukupuoli, siviilisääty ja sosiaalinen asema. Dynaaminen profiili sen sijaan päivittyy käyttäjän tarpeiden ja kiinnostuksen kohteiden muuttuessa. Näistä dynaaminen profiili on toivottavampi lähestymistapa, jotta pystyttäisiin paremmin täyttämään käyttäjän muuttuvat toiveet ja vaatimukset. Staattisen profiilin etuna on sen ylläpidon helppous.

Toinen tapa jaotella profilointimenetelmiä on puhua *tietoon pohjautuvista* tai *käyttäytymiseen pohjautuvista* menetelmistä (Middleton, Shadbolt ja De Roure 2004). Tämä on käytännössä vain hieman edellä mainitusta ”dynaaminen tai staattinen”—jaottelusta eroava lähestymistapa aiheeseen. Tietoon pohjautuvat menetelmät aloittavat yleensä staattisen tiedon keräämisellä esimerkiksi kyselylomakkeilla ja liittävät tämän jälkeen dynaamisesti käyttäjät sopiviin malleihin. Käytökseen pohjaavat menetelmät taas tutkivat nimensä mukaisesti käyttäjän käytöstä esimerkiksi selaushistorian perusteella ja muodostavat samalla dynaamisesti profiilia.

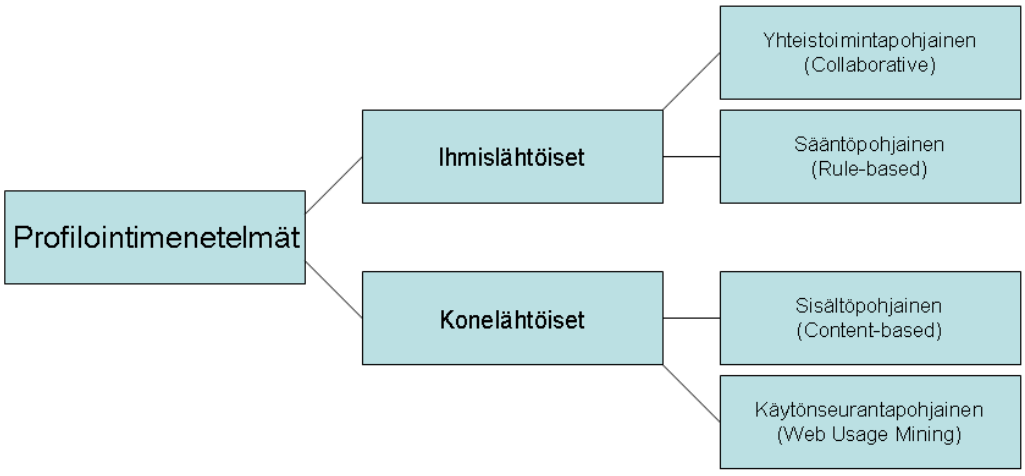
Edelliset jakotavat ovat kuitenkin hieman liian yleisiä hyödynnettäväksi menetelmien tutkimiseen. Tästä syystä seuraavaksi esitellään Eirinakin ja Vazirgianniksen (2003) esittämä erottelutapa, joka kertoo huomattavasti hyödyllisempää tietoa profilointimenetelmien toiminnasta kuin edellä esitellyt jaottelut. Tätä uutta jakoa käytetään pohjana tästä eteenpäin.

Kaksi pääluokitusta profilointimenetelmien erotteluun ovat käyttäjän tai käyttäjien itsensä määrittämät eli ihmislähtöiset profiilit (kohdassa 3.1.1 ja 3.1.2) sekä automaattisesti muodostetut eli konelähtöiset profiilit (kohdassa 3.1.3 ja 3.1.4) (Eirinaki ja Vazirgiannis 2003). Näiden luokkien sisältämiä profilointitapoja voidaan jossain määrin myös yhdistellä keskenään (kohdassa 3.1.5).

Ihmislähtöiset profiilit ovat yksittäisen käyttäjän tai käyttäjäryhmän tietoisesti määrittämä. Profiilin määrittäminen voi näin ollen tapahtua joko antamalla käyttäjälle lista, josta hän voi valita kiinnostavimmat aihealueet tai käyttämällä usean käyttäjän luomaa pisteytysjärjestelmää. Monesti tällaisessa tapauksessa käyttäjä rekisteröityy kauppapaikkaan, eikä tällöin käytetä evästepohjaista käyttäjän tunnistusmenetelmää (lisää evästeistä kohdassa 2.3) (esimerkiksi Amazon.com ja MovieLens).

Konelähtöiset profiilit perustuvat käyttäjän toimien seurantaan ja niihin reagointiin automaattisesti. Konelähtöisten menetelmien etuna on käyttäjän kuormittamisen vähentyminen sekä profiilin automaattinen päivittyminen. Jälkimmäinen näistä eli käytönseurantapohjainen profilointi vaikuttaisi Eirinakin ja Vazirgianniksen (2003) tutkimusten mukaan hyvältä profilointitavalta. Menetelmiin liittyvistä ongelmista kerrotaan tarkemmin jokaisen menetelmän kohdalla.

Profilointimenetelmien päätyypit esitetään kuvassa 4. Kuvassa profilointimenetelmät jakaantuvat aluksi ihmislähtöisiin ja konelähtöisiin menetelmiin. Ihmislähtöiset profilointimenetelmät jakaantuvat edelleen yhteistoiminta- ja sääntöpohjaisiin menetelmiin. Konelähtöiset profilointimenetelmät taas jakaantuvat sisältö- ja käytönseurantapohjaisiin menetelmiin.



Kuva 4. Profilointimenetelmien päätyypit.

Seuraavaksi tarkastellaan näitä neljää profiloinnin menetelmää yksityiskohtaisemmin. Jokaisen alaluvun alla esitellään ensin menetelmän perusominaisuudet ja tämän jälkeen käsitellään näihin liittyviä ongelmia. Viimeisessä alaluvussa (3.1.5) käsitellään yhdistettyjä menetelmiä, jotka sisältävät ominaisuuksia kahdesta tai useammasta profiloinnin menetelmästä.

3.1.1 Yhteistoimintapohjainen profilointi

Yhteistoimintapohjaisessa (engl. collaborative) profiloinnissa käyttäjiä pyydetään pisteyttämään tutkimiaan objekteja (esimerkiksi artikkeleita). Tämän jälkeen tutkitaan käyttäjiä, jotka ovat pisteyttäneet eri objekteja samalla tavalla. Saadun tuloksen pohjalta suositellaan etsijälle esimerkiksi artikkeleita, jotka ovat kiinnostaneet samalla tapaa objekteja arvostelleita käyttäjiä. (Eirinaki ja Vazirgiannis 2003)

Ongelmana menetelmässä on ennen kaikkea käyttäjän jatkuva kuormittaminen. Useasti tämäntyyppisissä järjestelmissä on huomattu, että käyttäjät eivät jaksa pisteyttää artikkeleita, jolloin profiilit eivät päivity ja tuloksen relevanttius käyttäjälle laskee (Carreira ym. 2004). Lisäongelmana on, että ennen kuin saadaan riittävän laaja pohja profiileille, saattavat suositukset olla jopa harhaanjohtavia.

3.1.2 Sääntöpohjainen profilointi

Sääntöpohjainen (engl. rule-based) profilointi perustuu yksittäisten käyttäjien pisteyttämiin listoihin. Nämä listat muodostuvat yleensä yleisen tason esimerkiksi demografisista, psykografisista tai maantieteellisistä tiedoista. Listat saattavat myös sisältää tarkemmin juuri kauppapaikkaan liittyviä tietoja. Yleisemmällä tasolla voidaan kysyä esimerkiksi tulotasoa, sosiaalista asemaa ja perhetilannetta. Kauppapaikkaan liittyvät listat taas keskittyvät tarkemmin juuri kyseiseen kauppapaikkaan liittyviin aiheisiin, esimerkiksi autoja myyvässä kaup-

papaikassa voidaan kysyä, että minkälaisista autoista käyttäjä tarkalleen ottaen on kiinnostunut. Tässä menetelmässä ei käytetä apuna muiden käyttäjien profiileita pisteytyksessä vaan luotetaan yksittäisten henkilöiden antamien tietojen tarkkuuteen (Eirinaki ja Vazirgiannis 2003).

Sääntöpohjaisen, aivan kuten yhteistoimintapohjaisen profilointitavankin ongelmana on käyttäjän kuormittaminen profiilin muodostamisen ja ylläpidon yhteydessä (Sugiyama, Hatano ja Yoshikawa 2004). Tällaisten profilointitapojen ongelmana on nähty se, että profiili ei pysy mukana käyttäjän mieltymysten muuttuessa ilman erillistä päivittämistä käyttäjän puolelta (Sugiyama ym. 2004, sekä Carreira ym. 2004).

Seuraavaksi esitellään kaksi erityyppistä konelähtöisen profiloinnin ratkaisua: sisältöpohjainen ja käytönseurantapohjainen profilointi.

3.1.3 Sisältöpohjainen profilointi

Sisältöpohjainen (engl. content-based) profilointi on konelähtöistä profilointia. Menetelmä perustuu yksittäisen käyttäjän aikaisemmin tekemiin valintoihin, joita järjestelmä seuraa. Näiden valintojen pohjalta järjestelmä suosittaa käyttäjälle häntä mahdollisesti kiinnostavia objekteja. Seurattavia asioita ovat esimerkiksi sivuilla navigointi ja siihen kulutettu aika tai vaikkapa tuotteet joita asiakas on ostanut. Menetelmässä hyödynnettävät tiedot kerätään yleensä vain yhdeltä kauppapaikalta. (Eirinaki ja Vazirgiannis 2003)

Sisältöpohjaisen menetelmän voisi sanoa olevan automaattinen versio edellä mainitusta käyttäjän itse määrittämästä yhteistoimintapohjaisesta menetelmästä. Ongelmana tällaisessa profiloinnissa on käyttäjälle tarjottavien objektien järkevä pisteytys. Käyttäjä on saattanut löytää tarvitsemansa, vaikka hän viipyisi objektin parissa vain muutaman sekunnin. Ei ole kuitenkaan syytä olettaa, että jos objekti on vain ladattu, on se myös ollut hyödyllinen ja käyttäjää tyydyttävä. Toinen ongelma on se, miten pisteytetään objektit, joita kukaan ei ole vielä tut-

kinut (Pazzani 1999). Tästä johtuen saattavat aloittamisvaiheen ongelmat nousta liian suuriksi.

3.1.4 Käytönseurantapohjainen profilointi

Käytönseurantapohjainen (engl. web usage mining) profilointi, josta saatetaan käyttää myös nimitystä www-käytön louhinta, on toinen ilman käyttäjän suoraa osallistumista määräytyvä profilointimenetelmä. Käytönseurantapohjainen profilointi on helppo sekoittaa sisältöpohjaisen profiloinnin kanssa, mutta nämä ovat kuitenkin toisistaan eroavia joiltakin olennaisilta piirteiltään. Tässä menetelmässä on tarkoituksena hyödyntää palvelimilla olevaa tietoa eli *www-lokitiedostoa* (engl. web log) käyttäjien toiminnan ymmärtämiseen. Selvänä erotuksena edelliseen menetelmään lähdetään käytönseurantapohjaisessa profiloinnissa muodostamaan profiilia lähtökohtana yksittäinen käyttäjä, eikä niinkään kauppapaikan suunnasta.

Menetelmän mukainen profiilin muodostus on kolmivaiheinen prosessi. Ensimmäiseksi prosessoidaan palvelimilla oleva käsittelemätön tieto, jotta pystyttäisiin erottamaan tiedon komponentit (muun muassa eri käyttäjät ja istunnot). Toisessa vaiheessa yritetään etsiä käsitellystä tiedosta kiinnostavia malleja. Tämä toteutetaan käyttäen erilaisia tilastollisia sekä tiedonsuodattamisen menetelmiä. Kolmannessa vaiheessa saatuja malleja analysoidaan tarkemmin, jotta pystyttäisiin luomaan profiileja käyttäjistä ja/tai käyttäjäryhmistä. (Eirinaki ja Vazirgiannis 2003)

Näissä kahdessa jälkimmäisessä konelähtöisesti muodostetussa profiilissa saattaa nousta ongelmaksi käyttäjän tunnistaminen: samaa konetta saattaa käyttää useampi käyttäjä. Tätä ongelmaa voidaan yrittää korjata manuaalisella sisäänkirjautumisella.

3.1.5 Yhdistetyt menetelmät

Jotta saataisiin mahdollisimman luotettava ja dynaaminen profiili, on mahdollista tarvittaessa yhdistellä näitä edellä esiteltyjä, toisistaan poikkeavia menetelmiä. Periaatteessa esitellyistä menetelmistä voidaan lähes kaikki, kokonaan tai vain joiltakin ominaisuuksiltaan yhdistää keskenään. On kuitenkin syytä pitää mielessä, että käyttäjää tulisi rasittaa mahdollisimman vähän, niin profiilin muodostus- kuin ylläpitovaiheessakin.

Valittaessa menetelmiä tulee huomioida menetelmien vaatima resurssientarve eli järjestelmän monimutkaisuus. Mitä useampi menetelmä on käytössä, sitä monimutkaisempi on järjestelmä ja sitä enemmän se vaatii resursseja palvelin-, ohjelmisto- sekä työntekijäpuolella. Monimutkaisuuden lisääminen ei välttämättä myöskään tarkoita, että järjestelmä olisi parempi, vaan se saattaa jopa olla esteenä toimivalle ja tehokkaalle profilointimenetelmälle.

Eräs tapa yhdistää menetelmiä on liittää toisiinsa ainakin joiltakin ominaisuuksiltaan konelähtöiset sekä ihmislähtöiset profiilinmuodostusmenetelmät. Tämä siksi, että saavutettaisiin ihmislähtöisesti määritetyn tiedon luotettavuus yhdistettynä konelähtöisesti toimivan järjestelmän käyttäjän kuormittamista vähentävään ominaisuuteen.

Kone- ja ihmislähtöisesti määritettyjen profiilien yhdistäminen voidaan toteuttaa esimerkiksi seuraavalla tavalla: kauppapaikkaan ensimmäistä kertaa tullessaan annetaan käyttäjälle mahdollisuus määrittää ihmislähtöisesti joitakin sivulle ominaisia ominaisuuksia esimerkiksi taulukon muodossa. Tämä kannattaa kuitenkin pitää vapaaehtoisena toimenpiteenä. Tämän jälkeen hoidetaan profiilin tarkennus ja päivitys konelähtöisten menetelmien avulla.

Tällaisen yhdistelmän avulla pystyttäisiin joiltain osin välttämään konelähtöisten menetelmien ongelmallinen alkuvaihe, jolloin profiilia ei ole lainkaan tai se

ei ole tarpeeksi tarkka. Käyttäjää ei tällä tavalla myöskään kuormitettaisi liikaa profiilin ylläpitovaiheessa.

Sugiyama ym. (2004) ehdottavat hakukoneiden hakutulosten personointiin profilointitapaa, joka yhdistää muunnelman yhteistoimintapohjaisesta profiloinnista (3.1.1) sekä käytönseurantapohjaisesta profiloinnista (3.1.4). Näin he yrittävät saavuttaa usean käyttäjän profiilien yhdistämisen hyödyn ja yhdistää sen tietokoneavusteiseen profiilinmuodostamiseen, jolloin saadaan käyttäjän työkuormaa vähennettyä. Tämä tapahtuu lisäämällä käyttäjän selaushistoriatietoihin usean käyttäjän profiileita vertailemalla saadut tulokset.

Sugiyaman ym. (2004) lisätavoitteena on muodostaa profiili, joka ymmärtää niin lyhyt- kuin pitkäaikaisiakin muutoksia käyttäjän tavoitteissa. Heidän on tarkoitus saavuttaa nämä kummatkin tavoitteet vaatimatta käyttäjältä mitään muita toimia kuin normaalin hakutulosten selailun (Sugiyama ym. 2004). Kuten muissakin konelähtöisesti muodostetuissa profilointitavoissa, on myös tässä menetelmässä ongelmana se, että tulokset alkavat tarkentua vasta kun järjestelmässä on ollut käyttäjiä jo jonkun aikaa. Jos käyttö on harvaa ja käyttötarkoitukset ristiriitaisia keskenään, kestää tarkentuminen luonnollisesti kauemmin

3.2 Tiedonsuodatusmenetelmät

Tiedonsuodatusmenetelmät ovat olennainen osa käyttäjien profilointia. Nämä suodatusmenetelmät saattavat olla niin päällekkäisiä itse profilointimenetelmän kanssa, että näitä kahta voi olla vaikeaa erottaa toisistaan. Tämä käy ilmi esimerkiksi siinä, että useassa lähteessä esitellään suodatusmenetelmien nimillä kokonaisia profilointimenetelmiä (Eirinaki ja Vazirgiannis 2003, sekä Yu, Tresp ja Yu 2004). Tiedonsuodatusmenetelmiä on tutkittu laajasti myös erikseen ilman varsinaista kytköstä profilointiin (muun muassa Belkin ja Croft 1992).

Erilaiset suodatusmenetelmät eroavat keskenään huomattavastikin. Menetelmät saattavat olla esimerkiksi kyselykieliä tai erilaisia algoritmeja, jotka tulkitsevat saatavilla olevaa tietoa omatoimisesti.

On tärkeää huomata, että yhden menetelmän hyödyntäminen ei välttämättä sulje pois muiden menetelmien käyttöä. Saattaa olla jopa tarpeellista ottaa käyttöön useampi näistä menetelmistä yhtäaikaisesti.

3.2.1 XML-QL

Kun käsitellään XML-muodossa olevaa tiedostoa, on ongelmana monesti suodattamattoman tiedon valtava määrä. Tätä ongelmaa ratkaisemaan on kehitetty muun muassa XML-QL (Deutsch, Fernandez, Florescu, Levy ja Suciu 1998). *XML-QL:lla* pystytään suorittamaan tietohakuja XML-muotoiseen tiedostoon ja muodostamaan tästä huomattavasti käyttökelpoisempaa informaatiota, kuin pohjalla ollut valtava määrä käsittelemätöntä tietoa.

XML-QL on syntaksiltaan läheisesti SQL-kieltä muistuttava, sisältäen SELECT-WHERE-rakenteen. Tällä rakenteella pystytään poimimaan XML-dokumentista esimerkiksi jonkun tietyn arvon omaavia elementtejä, kuten vaikkapa jonkun tietyn käyttäjän tietyn kauppapaikan vierailukerrat. (Deutsch, Fernandez, Florescu, Levy ja Suciu 1998)

Tällä tavalla saatu tieto ei välttämättä kuitenkaan ole riittävän tarkkaa tai suodatettua profiloinnin tarpeeseen, joten asiaa on pyritty lähestymään myös toisesta näkökulmasta (Xie ja Phoha, 2001). Näin ollen on kehitetty muun muassa algoritmeja jotka yrittävät muodostaa metatietoa profiilin tarjoamasta raakatiedosta. Seuraavassa kappaleessa esitellään näistä algoritmeista kaksi hiukan toisistaan poikkeavaa näkökulmaa. Nämä näkökulmat eivät eroa toisistaan lähtökohdiltaan, eivätkä lopputulokseltaan, vaan sen suhteen kuinka lopputulokseen päästään.

3.2.2 Algoritmit

Tiedon suodattamiseen käytettäviä algoritmeja on useita erilaisia. Yhteistä näille kaikille on tavoite sisäänrakennettuun älykkyyteen tietoa suodatettaessa.

Tässä esitellään kaksi erilaista algoritmia, jotka yrittävät ennustaa todennäköisyyksiä. Nämä todennäköisyydet saattavat esimerkiksi yrittää ennustaa miten tietynlainen käyttäjä tulee toimimaan tietynlaisessa tilanteessa seuraavaksi. Nämä kaksi esiteltävää ovat Bayesin teoreemaan pohjautuvat algoritmit, sekä Olettamusfunktio. Ensimmäisen algoritmin muunnelmia näkee paljon tiedonsuodatusta käsittelevissä artikkeleissa (Wong ja Butz 2000). Toinen algoritmi taas perustuu Dempster-Shaferin teoriaan (Yu ym. 2004).

Bayesin teoreemaan perustuvia menetelmiä käytetään yleisesti todennäköisyyksien laskemisessa. Bayes oli matemaatikko joka tutki todennäköisyyksiä. Hänen kehittämänsä teoreeman voisi kuvata olevan käytännönläheinen lähestymistapa todennäköisyyden laskemiseen. (Wong ja Butz 2000)

Teoreemaan pohjautuen on kehitetty useita eri sovelluksia myös tiedonsuodattamiseen www-profiloinnissa. Teoreeman sovellukset yrittävät etsiä malleja erilaisista tietolähteistä. Käyttötarkoituksena voi olla vaikkapa dokumenttien tai käyttäjien jaottelu. Bayesin teoreemaan pohjautuvat menetelmät voidaan jaotella vielä hierarkkisiin sekä ei-hierarkkisiin menetelmiin. (Yu ym. 2004)

Xie ja Phoha (2001) ehdottavat käytettäväksi O*lettamusfunktiota* (engl. belief function) tiedonsuodatukseen profiilin muodostamissa ja ylläpidossa. Funktio perustuu Dempster-Shaferin teoriaan. Tätä teoriaa kutsutaan myös Uskomusfunktioiksi. Dempster-Shaferin teoria on yleistys Bayesin teoriasta. Olettamusfunktiot eroavat Bayesin teoriasta siinä, että ne eivät vaadi selkeitä numeerisia arvoja jokaiselle mahdollisuudelle, vaan arvot voidaan perustaa olettamuksiin.

Xien ja Prohan (2004) mallissa käytetään Olettamusfunktiota luokittelemaan käyttäjiä erilaisiin luokkiin. Näitä luokkia voidaan käyttää normaaleissa profi-

loinnin ja personoinnin kohteissa kuten esimerkiksi www-sivujen personoinneissa, hakutulosten muokkauksessa, sekä mainosten kohdistamisessa.

Sekä Bayesin, että Dempster-Shaferin teoria (eli Olettamusfunktio) on kehitetty todennäköisyyksien laskentaan, kun tarvitaan arvioita lopputuloksesta ilman selkeitä lähtökohtia. Teorioiden voidaan sanoa olevan vaihtoehtoisia lähestymistapoja monilähteisen tiedon yhdistämiseen. Teorioiden käytännöllisyydestä ja siitä kumpi on paremmin tarkoitukseen sopiva, kiistellään edelleen. (Braun 2000)

Tässä luvussa luotiin kirjallisuuskatsaus olemassa oleviin profilointimenetelmiin. Havaittiin, että on olemassa kahta päälajia profilointia: ihmis- sekä konelähtöistä. Nämä jakaantuvat vielä neljään osaan jotka ovat: yhteistoimintapohjainen, sääntöpohjainen, sisältöpohjainen ja käytönseurantapohjainen profilointi. Tämän lisäksi tutkittiin tiedonsuodatusmenetelmiä, joita tunnistettiin XML-QL sekä kaksi erilaista algoritmia.. Seuraavassa luvussa luodaan näiden tulosten pohjalta viitekehys profilointimenetelmien arviointia varten.

4 PROFILOINTIMENETELMIEN OMINAISUUKSIEN VIITEKEHYS

Tässä luvussa muodostetaan viitekehys profilointimenetelmien arviointia varten. Pohjana tälle Laverin viitekehykselle käytetään luvuissa kaksi ja kolme saatuja tuloksia. Luku vastaa Nunamakerin, Chenin ja Purdinin (1990) esittämässä mallissa (katso 1.3) kohtaa teorian rakentaminen. Teorian rakentamisessa konstruoidaan esimerkiksi teoreettinen viitekehys (Nunamaker ym. 1990, 94).

Profilointimenetelmien jaottelu ei ole helppoa ja erilaisia jaottelumalleja on useita (esimerkiksi Wu, Tremaine, Instone ja Turoff 2002 sekä Eirinaki ja Vazirgiannis 2003). Menetelmät, jotka vaikuttavat aluksi hyvinkin erilaisilta, saattavat paljastua täysin samanlaisiksi erilaisista termeistä huolimatta. Tämä sekavuus määrityksissä vaikeuttaa huomattavasti niin profiloinnin yleistä tutkimista kuin myös sopivan menetelmän valintaa käyttäjien profilointia harkittaessa.

Tässä luvussa hahmotetaan profilointimenetelmistä niitä määrittäviä ominaisuuksia mahdollisimman selkeiksi kokonaisuuksiksi. Näistä ominaisuuksista muodostetaan viitekehys, jonka avulla pystytään hahmottamaan eri menetelmien erilaiset osat. Tämän Laverin viitekehyksen kolme ensimmäistä osaa ovat profiilin elinkaaressa peräkkäisiä, mutta suurin osa ominaisuuksista on mukana koko profiilin elinkaaren ajan.

Muodostettava viitekehys pyrkii olemaan vertailukehikko www-ympäristössä käytössä oleville profilointimenetelmille. Kehyksellä voidaan tutkia jo käytössä olevien menetelmien tai järjestelmien ominaisuuksia tai sitä voidaan hyödyntää apuna valittaessa uutta profilointijärjestelmää. Kehystä voidaan myös käyttää hahmottamaan, minkälaisia ominaisuuksia erilaisissa profiloinnin menetelmissä tai järjestelmissä voi olla. Viitekehystä käytetään vertailemalla profilointimenetelmän tai –järjestelmän ominaisuuksia viitekehyksessä mainittuihin ominaisuuksiin.

Jokainen tämän luvun alakappale käsittelee yhtä profilointimenetelmien ominaisuusjoukkoa tarkemmin. Profilointimenetelmän ei välttämättä tarvitse sisältää kaikkia tässä käsiteltyjä ominaisuuksia, mutta useimmiten se kuitenkin sisältää useamman kuin yhden mainituista ominaisuuksista.

Jokaisessa alaluvussa on kuva, jossa näkyvät vaiheessa tunnistetut ominaisuudet. Näistä osista muodostuu profilointimenetelmien arvioinnin viitekehys. Tämä Laverin viitekehys löytyy kokonaisuudessaan liitteestä A.

4.1 Käyttäjän tunnistus

Profiilin elinkaari alkaa käyttäjän tunnistamisella. Käyttäjän tunnistus on ratkaiseva tekijä niin profiilin hyödyntämisen kuin myös sen yleisen toimivuuden kannalta. Profiilia on mahdotonta rakentaa, jos käyttäjää ei pystytä tunnistamaan virheettömästi. Tämän työn kohdassa 2.3 on esitelty käyttäjän tunnistamiseen liittyviä asioita tarkemmin.

Käyttäjän tunnistamiselle on kolme eri vaihtoehtoa: selaimen eväste, sisäänkirjautuminen ja kolmas osapuoli. Evästeen ja sisäänkirjautumisen tapauksissa itse profiili sijaitsee kauppapaikan tai yrityksen palvelimella. Komannen osapuolen tapauksessa koko profiili sijaitsee tämän ulkopuolisen tahon hallussa. Tämä kolmas osapuoli voi olla esimerkiksi aiemmin mainittu ”luotettu taho”. Kuvassa 5 esitellään käyttäjän tunnistuksen ominaisuudet. Seuraavaksi määritellään kuvan sisältämät termit.

Käyttäjän tunnistus	
	Eväste
	Sisäänkirjautuminen
	Profiili kolmannen osapuolen hallussa

Kuva 5. Käyttäjän tunnistus

Käyttäjä voidaan tunnistaa *evästeen* avulla. Eväste sijaitsee käyttäjän koneella siinä selaimessa, jota käytettiin kun vierailtiin kauppapaikalla. Tällainen säilytystapa on ongelmallinen, koska profiili ei siirry edes saman koneen eri selaimien välillä. Vielä vaikeampaa on käyttäjän tunnistus, jos käytetään eri päätelaitetta kuten kännykkää kotikoneen sijaan. Käyttäjä tai erinäiset ohjelmat saattavat myös poistaa evästeen koneelta. Jos käytetään sekä samaa selainta että konetta, eikä evästettä ole syystä tai toisesta poistettu, on käyttäjän tunnistaminen kätevää, koska evästeen ansiosta hänet voidaan yksilöidä aukottomasti kauppapaikassa. Käytännössä eväste ei sisällä varsinaista profiilia vaan vain pienen määrän tietoa, jolla voidaan tunnistaa asiakas. Tästä huolimatta evästeen poistolla on yhtä vakavat seuraukset profiilin toimivuudelle, kuin jos eväste sisältäisi koko profiilin. (Cingil 2002)

Käyttäjä voidaan *tunnistaa myös sisäänkirjautumisen avulla*. Tämä on profiilin hyödyntämisen ja siirrettävyyden kannalta parempi tilanne kuin evästeenä säilytettävä tunniste. Käyttäjien tunnistamisen muilla tavoin, esimerkiksi käyttäjän koneen IP-osoitteen avulla ei ole luotettavaa IP-osoitteiden ja koneen käyttäjien muuttumisen takia. Sisäänkirjautuminen onkin luotettavuutensa ja turvallisuutensa takia hyvä tapa tunnistaa käyttäjä. (Cingil 2002)

Viimeinen vaihtoehto on, että koko *profiili sijaitsee jonkun kolmannen osapuolen hallussa*. Tämä osapuoli voi olla esimerkiksi ”luotettu taho” (katso 2.3.). ”Luotettu taho” on riippumaton taho, jonka palvelimella sijaitsee käyttäjän profiili. Näin ollen kauppapaikat pystyvät, saamaan tietoonsa käyttäjän profiilin, jos käyttäjä antaa siihen luvan. Käytettäessä luotettua tahoa saattaa olla mahdollista välttyä kylmäkäynnistys ongelmalta, olettaen, että jo olemassa olevasta profiilista on hyötyä kauppapaikalle. (Cingil 2002)

Kylmäkäynnistyksellä (engl. cold start) tarkoitetaan ongelmaa, joka on yleensä konelähtöisesti muodostettavien profilointimenetelmien muodostusvaiheessa,

kun hallussa ei vielä ole tarvittavaa tietokantaa luotettavien profiilien koneelliseen muodostamiseen (Middleton 2003, 65).

4.2 Profiilin muodostusvaihe

Käyttäjän tunnistuksen jälkeen aletaan kerätä profiiliin varsinaista sisältöä. Tässä profiilin muodostusvaiheessa asiakas voidaan karkottaa kaupasta esimerkiksi liian pitkillä ja vaivalloisesti täytettävillä listoilla tai pelkästään sillä, että vaaditaan liian yksityiskohtaisia tai henkilökohtaisia tietoja. Mikä muodostusvaiheessa käytettävä menetelmä sitten onkin, täytyy sen tuottaa oikeaa tietoa järjestelmälle. Ilman tätä oikeaa tietoa järjestelmä saattaa pahimmillaan jopa häiritä käyttäjien toimia virheelliseen profiiliin pohjautuen. Nämä haitat voivat olla vääristyneitä hakutuloksia tai suosituksia tuotteista, jotka eivät asiakasta kiinnosta.

Kuvassa 6 näkyvät ominaisuudet tai ominaisuusjoukot, joita tässä vaiheessa on tunnistettu. Kuvassa näkyvät ominaisuudet ovat muodostusvaiheen pääjako. Seuraavaksi määritellään kuvan sisältämät termit.

Profiilin muodostusvaihe	
	Valmiiden tietokantojen tuonti
	Ihmislähtöinen profilointi
	Konelähtöinen profilointi

Kuva 6. Profiilin muodostusvaiheen ominaisuudet

Vaikka Eirinakin ja Vazirgianniksen (2003) profilointimenetelmien jaosta sisältöpohjainen sekä käytönseurantapohjainen profilointi kuuluvat konelähtöisiin profilointimenetelmiin käsitellään nämä tässä. Tämä johtuu siitä, että näistä

profilointimenetelmistä saadaan ominaisuudeksi *valmiiden tietokantojen tuonti* ja tämä ominaisuus ei kuulu ihmis- eikä konelähtöiseen profilointiin.

Profilointijärjestelmään saattaa olla mahdollista *tuoda valmiita tietokantoja*. Tämä mahdollisuus voi helpottaa kylmäkäynnistysongelmia kauppapaikassa. Tietokantojen tuomisessa ongelmia saattaa seurata tietokantojen tai niiden sisältämien profiilien yhteensopimattomuudesta.

Edellisessä luvussa mainittiin profilointimenetelmien kaksi erillistä pääosaa. Nämä olivat ihmislähtöinen ja konelähtöinen profilointi. Nämä kaksi termiä otetaan myös tässä lähtökohdiksi. Näiden termien sisältö on sama kuin aikaisemminkin: ihmislähtöisessä määrityksessä käyttäjä on aktiivisesti ja tietoisesti mukana profiilin muodostamisessa, kun taas konelähtöisessä määrityksessä käyttäjä ei ole välttämättä lainkaan tietoinen profilointitapahtumasta, eikä se ainakaan vaadi häneltä erillisiä toimenpiteitä. Ihmislähtöisen ja konelähtöisen profiloinnin ominaisuuksia tarkennetaan seuraavissa alaluvuissa (4.2.1 ja 4.2.2)

4.2.1 Ihmislähtöinen profilointi

Ihmislähtöinen (joissakin yhteyksissä *eksplisiittinen*) profilointi kuvastaa käyttäjän avulla määritettyä profiilia. Tällaisessa tilanteessa käyttäjä itse on aktiivisena osana muodostamassa omaa profiiliaan esimerkiksi täyttämällä erilaisia listoja tai arvostelemalla objekteja.

Ihmislähtöisien profilointimenetelmien ongelmana nähdään se, että profiiliin vaadittavien tietojen kerääminen kuormittaa käyttäjää liiaksi (Sugiyama, Hatano ja Yoshikawa 2004). Kuten aiemmin jo todettiin, voi tämänkaltainen kuormittaminen olla syynä siihen, että kauppapaikan profilointi hylätään käyttäjän puolelta täysin. Eräs erittäin hyvä puoli tällaiseen profilointimenetelmään kuitenkin sisältyy: kylmäkäynnistysongelma poistuu lähes kokonaan.

Eirinakin ja Vazirgianniksen (2003) profilointimenetelmien jaosta ihmisläheiseen profilointiin kuuluvat sääntöpohjainen sekä yhteistoimintapohjainen profilointi.

Sääntöpohjaisissa profilointimenetelmissä annetaan käyttäjälle suuri vastuu ja myös oletetaan, että hän kertoo totuudenmukaisia asioita itsestään. Sääntöpohjaiseen profilointimenetelmään sisältyy yleisesti erilaisten listojen täyttämistä. Nämä listat voivat sisältää erilaisia kysymyksiä liittyen joko käyttäjän yleisiin ominaisuuksiin tai esimerkiksi vain kyseiseen kauppapaikkaan liittyviin tietoihin (Eirinaki ja Vazirgiannis 2003)

Sääntöpohjaisesta profilointimenetelmästä saadaan näin ollen menetelmien ominaisuuksiksi käyttäjän täyttämät yleisten tietojen listat sekä käyttäjän täyttämät kauppapaikkaan liittyvät listat.

Yhteistoimintapohjaisissa profilointimenetelmissä useampi käyttäjä pisteyttää ja/tai arvostelee kauppapaikassa tai sivustolla olevia objekteja, kuten uutisartikkeleita tai tuotteita. Tämän jälkeen luokitellaan samaan tapaan objekteja arvostelleita asiakkaita ja tuloksen perusteella suositellaan heille objekteja, joita saman maun omaavat asiakkaat ovat tutkineet tai hankkineet. (Eirinaki ja Vazirgiannis 2003)

Yhteistoimintapohjaiset profilointimenetelmät eroavat muista ihmisläheisen profiloinnin tavoista siten, että ne kärsivät kylmäkäynnistysongelmasta. Kylmäkäynnistyshän on ongelmana yleensä lähinnä konelähtöisissä profilointimenetelmissä. Ongelma muodostuu jälleen siitä, että järjestelmää käynnistettäessä ei ole olemassa vielä tietokantaa, jonka perusteella objekteja ja käyttäjiä voisi luokitella (Carreira, Crato, Gonçalves ja Jorge 2004). Tätä ongelmaa on kuitenkin mahdollista vähentää esittämälle käyttäjälle heti hänen rekisteröityessään kauppapaikkaan hyvin valittuja objekteja, jotka auttavat käyttäjän aseman hahmottamisessa suhteessa muihin (Rashid, Albert, Cosley, Lam, McNee, Konstan ja Riedl 2002). Kuten sanottua, näytettävien objektien täytyy olla hyvin va-

littuja, jotta ne todella auttavat profiilinmuodostusta. Esimerkiksi kysymällä ruuanlaittosivustolla ”pidätkö pirtelöstä?” ei saada kovin tarkkaa erottelua käyttäjien kesken, koska suurin osa käyttäjistä varmasti pitää pirtelöstä. Objektin tulisi kuitenkin olla tarpeeksi suosittu, jotta pystytään vertailemaan käyttäjiä keskenään. Näin ollen optimaalinen objekti olisi kaikkien käyttäjien arvostelema ja pisteytetty mahdollisimman suurella hajonnalla (Rashid ym. 2002). Tämäkään menetelmä ei auta aivan ensimmäisten käyttäjien kohdalla, jolloin käytössä ei ole vielä tietokantaa käyttäjistä tai objekteista. Tällöin ainut vaihtoehto on saada käyttöön joku jo olemassa oleva tietokanta, jonka perusteella voidaan valita ensimmäisenä käyttäjille näkyvät objektit.

Yhteistoimintapohjaisista profilointimenetelmistä saadaan ominaisuuksiksi käyttäjän tekemät objektien arvostelut, käyttäjien luokittelu ryhmiin sekä käyttäjien keskenään vertailu. Näistä kaksi jälempää ovat profiilinhyödyntämisen ominaisuuksia ja käsitellään kohdassa 4.4.

Edellä tunnistetut ihmislähtöisen profiloinnin ominaisuudet näkyvät kuvassa 7. Seuraavaksi määritellään kuvan sisältämät termit.

Ihmislähtöinen profilointi	
	Yleisten tietojen listat
	Kauppapaikkaan liittyvät listat
	Objektien arvostelu

Kuva 7. Ihmislähtöisen profiloinnin ominaisuudet

Yleisten tietojen listat sisältävät yleistä tietoa käyttäjän asemasta yhteiskunnassa. Listoilla saatetaan kysyä esimerkiksi seuraavia asioita: sukupuoli, tulotaso, per-

hetilanne tai asuinpaikka. Näitä listoja ei tarvitse niiden yleisen luonteen takia välttämättä määritellä jokaiselle kauppapaikalle erikseen.

Kauppapaikkaan liittyvät listat ovat kauppapaikan itsensä määrittämiä listoja, jotka liittyvät läheisesti sen aihepiiriin. Tällaiset listat sisältävät kysymyksiä, jotka voivat esimerkiksi autoiluun liittyvässä kauppapaikassa olla asiakkaan auton merkki, malli, sekä vuosimalli.

Käyttäjän tekemät objektien arvostelut koskevat objekteja, jotka sijaitsevat kauppapaikassa tai sivustolla. Tällaisia objekteja voivat olla esimerkiksi kauppapaikassa myytävät tuotteet tai sivustolla sijaitsevat artikkelit.

4.2.2 Konelähtöinen profilointi

Konelähtöisissä (joissakin yhteyksissä *implisiittinen*) profilointimenetelmissä profiili muodostetaan seuraamalla koneellisesti käyttäjän toimia. Tällaisessa tilanteessa seurataan hyvinkin huomaamattomasti esimerkiksi käyttäjän navigointia kauppapaikassa. Seuranta tapahtuu täysin koneellisesti.

Konelähtöisien profilointimenetelmien ongelmana on kylmäkäynnistys. Tämä johtuu siitä, että tällaisessa konelähtöisesti muodostetussa profiilissa ei ole alussa vielä lainkaan tietoa. Joissakin järjestelmissä tätä voidaan yrittää korjata tuomalla jo olemassa olevista järjestelmistä valmiita tietokantoja. Nämä eivät kuitenkaan välttämättä ole toimiva ratkaisu kahdesta pääsyystä. Ensinnäkin käyttäjien erottaminen tällä tavalla toisistaan on lähes mahdotonta ilman yhteistä profiilitietokantaa. Toiseksi, vaikka käyttäjät saataisiinkin tunnistettua, saattavat heidän tarpeensa olla erilaiset eri kauppapaikkojen kesken. Näin ollen minimivaatimuksena ominaisuudelle onkin samaan alaan keskittyvät kauppapaikat. (Middleton 2003, 65)

Eirinakin ja Vazirgianniksen (2003) profilointimenetelmien jaosta sisältöpohjainen sekä käytönseurantapohjainen profilointi kuuluvat konelähtöisiin profiloin-

timenetelmiin. Profilointimenetelmien ominaisuudeksi näistä saadaan mahdollisuus tuoda järjestelmään valmiita profiilitietokantoja.

Sisältöpohjaisessa määrityksessä seurataan käyttäjän toimia yhdessä kauppapaikassa. Näihin toimiin voi kuulua esimerkiksi yksittäisellä kauppapaikan sivulla viipymisen aika, sivujen väliset navigoinnit ja ostetut tuotteet. Kuten aiemmin mainittiin, suurin ongelma menetelmässä on miten käyttäjän toimia arvotetaan. Yksittäinen sivu on voinut olla käyttäjälle täysin hyödyllinen, vaikka hän olisi viipynyt sillä vain muutaman sekunnin. Toisaalta pitkäkään viipymisaika ei tarkoita, että sivu olisi välttämättä täyttänyt käyttäjän tarpeita. (Eirinaki ja Vazirgiannis 2003)

Käytönseurantapohjaisissa profilointimenetelmissä, aivan kuten sisältöpohjaisissakin profilointimenetelmissä, seurataan käyttäjän toimia. Käytönseurantapohjaisissa menetelmissä analysoidaan www-lokitiedostoa, joka syntyy, kun käyttäjä selailee www-sivuja. Tiedostoa analysoimalla yritetään löytää malleja, joiden perusteella pystytään muodostamaan profiileja.

Sisältöpohjaisista profilointimenetelmistä saadaan ominaisuuksiksi sivujen käyttämisen seuranta sekä ostojen seuranta. Käytönseurantapohjaisista profilointimenetelmistä saadaan ominaisuudeksi www-lokitiedostojen analysointi. Nämä ominaisuudet löytyvät kuvasta 8. Seuraavaksi määritellään kuvan sisältämät termit.

Konelähtöinen profilointi	
	Sivujen käytön seuranta
	Hankintojen seuranta
	Www-lokitiedostojen analysointi

Kuva 8. Konelähtöisen profiloinnin ominaisuudet

Sivujen käyttämisen seurannassa seurataan eri sivujen liikennettä. Seurattavia asioita ovat esimerkiksi kuinka kauan sivulla viivytään, mistä sivulle on päädytty ja mihin sieltä jatketaan. Sivuston sivut on syytä luokitella niiden tärkeysjärjestyksen mukaan, esimerkiksi sisällys- ja navigointisivut erikseen. Menetelmän tehokkuus kärsii ilman tällaista luokittelua.

Hankintojen seurannassa seurataan, miten kauppapaikan tuotteita on ostettu ja liitetään nämä tiedot ostajien profiileihin. Kuinka näitä profiileita hyödynnetään jatkossa, on kauppapaikkakohtaista.

Www-lokitiedostojen analysoinnissa tutkitaan palvelimilla olevaa tietoa. Tiedosta pyritään tällä tavoin muodostamaan käsiteltyä informaatiota, josta taas pystytään jatkokehittämään joko käyttäjien profiileita tai muuten ymmärtämään asiakkaiden toimia tarkemmin. Tähän menetelmään voi sisältyä joku tiedonsuodatuksenmenetelmä, jonka avulla voidaan analysoida lokitiedostojen sisältämät todella suuret määrät käsittelemätöntä tietoa.

4.3 Profiilin ylläpitovaihe

Profiilin muodostusvaihe on luultavasti tärkein yksittäinen osa profiilin elinkaaressa, mutta ei ole myöskään syytä vähätellä ylläpitovaiheen tärkeyttä. Ylläpitovaiheessa profiilia tarkennetaan vastaamaan paremmin profiloidun henkilön ominaisuuksia tai muuttuneita tarpeita. Kuvassa 9 näkyvät vaiheen ominaisuudet. Seuraavaksi selitetään nämä ominaisuudet.

Ylläpitovaihe	
	Ihmislähtöinen profilointi
	Konelähtöinen profilointi

Kuva 9. Profiilin ylläpitovaiheen ominaisuudet

Ylläpitovaiheessa on olemassa jo joku profiili, jota päivitetään. Tähän vaiheeseen sisältyvät ominaisuudet joko hoitavat ylläpidon tai auttavat siinä. Käytännössä ylläpitovaiheeseen sisältyy kaikki samat profilointimenetelmien ominaisuudet kuin muodostusvaiheeseenkin, lukuun ottamatta mahdollisuutta tuoda valmiita tietokantoja (kuva 9). Tästä johtuen ihmisläheisen profiloinnin (katso 4.2.1) ja konelähtöisen profiloinnin (katso 4.2.2) menetelmiä ei tässä kohdassa uudestaan kuvata. Jos käytetään vain yhtä menetelmää molemmissa vaiheissa, merkitään tällöin sama menetelmä niin muodostamis- kuin ylläpitovaiheeseenkin.

4.3.1 Vaikeasti luokiteltavat profilointimenetelmät

Osasta menetelmiä on mahdotonta erottaa toisistaan muodostusvaihetta ja päivitysvaihetta. Tässä esiteltyjen ominaisuuksien perusteella profilointimenetelmät, joista on vaikeaa erottaa näitä vaiheita toisistaan, näyttäisivät olevan konelähtöisen profiloinnin menetelmiä. Kuten aiemmin jo mainittiin, on näissä menetelmissä kylmäkäynnistysongelmia eli alussa menetelmässä ei ole riittävästi tai lainkaan tietoa. Näin ollen muodostusvaiheen profilointi on käytännössä jo päivitysvaihetta.

Kohdassa 4.2 tutkituista profilointimenetelmistä näihin vaikeasti luokiteltaviin kuuluvat konelähtöiset sisältöpohjainen sekä käytönseurantapohjainen profilointimenetelmä. Näiden lisäksi vaikeasti luokiteltava on myös ihmislähtöinen yhteistoimintapohjainen menetelmä, joka sisältää samoin kylmäkäynnistyksen ongelman. Ominaisuuksista vaikeasti luokiteltaviin kuuluvat seuraavat: käyttäjällä mahdollisuus arvostella objekteja, käyttäjien luokittelu ryhmiin, sivujen käyttämisen seuranta, ostojen seuranta ja www-lokitiedostojen analysointi.

Huolimatta siitä, että näistä mainituista menetelmistä on niiden muodostus- ja ylläpitovaihetta vaikea erottaa toisistaan, on nämä menetelmät silti syytä sisäl-

lyttää molempiin vaiheisiin. Menetelmiä yhdistettäessä voidaan jotain toista menetelmää käyttää esimerkiksi muodostusvaiheessa ja tällaista vaikeasti määritettävää käytetään vain ylläpitovaiheessa. Esimerkiksi käyttäjän täyttämiä käyttäjän yleisten tietojen listaa käytetään muodostamisvaiheessa ja vasta ylläpitovaiheessa otetaan mukaan esimerkiksi sivujen käyttämisen seuranta.

4.4 Profilointimenetelmän tiedonsuodatus

Profilointimenetelmän tiedonsuodatukseen voidaan käyttää useita hyvinkin erilaisia menetelmiä. Joissakin tapauksissa nämä menetelmät ovat myös pääasiallinen tapa muodostaa käyttäjän profiili. Tässä työssä keskitytään kolmeen kohdassa 3.4 esiteltyyn menetelmään tai niiden muunnelmaan. Nämä menetelmät ovat XML-QL (katso 3.1), Bayesin teoreeman johdannaiset (katso 3.2.1), sekä Olettamusfunktion johdannaiset (katso 3.2.2) (kuva 10). Kahta jäljempää näistä voidaan käyttää joissakin tapauksessa yksinään ilman muita menetelmiä profiilinmuodostusprosessissa. Nämä menetelmät toimivat myös sellaisenaan ominaisuuksina viitekehystä varten. Seuraavaksi määritellään kuvan sisältämät termit.

Profilointimenetelmän tiedonsuodatus	
	XML-QL
	Bayesin teoreeman johdannaiset
	Olettamusfunktion johdannaiset

Kuva 10. Profilointimenetelmän tiedonsuodatus

Profiilin suodatusmenetelmän ollessa *"XML-QL"*, käyttää profilointimenetelmä tällöin tiedonsuodatukseen XML-QL-kyselykieltä (katso 3.2). XML-QL ei yksi-

nään sovellu profiilinmuodostukseen vaan se vaatii kumppanikseen muita menetelmiä. Profiilin suodatusmenetelmän ollessa *"Bayesin teoreeman johdannaiset"*, käyttää profilointimenetelmä tällöin jotain Bayesin teoreeman johdannaisalgoritmia (katso 3.2.1) joko tiedonsuodatukseen tai koko profiilinmuodostus prosessiin. Ja viimeisenä, jos profiilin suodatusmenetelmäksi tulee *"Olettamusfunktion johdannaiset"*, käyttää profilointimenetelmä tällöin jotain olettamusfunktion johdannaisalgoritmia (katso 3.2.2) joko tiedonsuodatukseen tai koko profiilinmuodostus prosessiin.

4.5 Profiilin hyödyntäminen

Pelkästään tietoa keräämällä ei ratkaista asiakasuskollisuuden tai asiakashallinnan ongelmia. Profiilin muodostus- ja ylläpitovaiheessa kerättyjä tietoja täytyy myös hyödyntää jollain tavalla, jotta niistä olisi jotain hyötyä. Tähän vaiheeseen kuuluvia ominaisuuksia ovat käyttäjien luokittelu ryhmiin, sekä vertailu käyttäjien kesken (kuva 11). Jälkimmäisestä ominaisuudesta on kaksi hiukan toisistaan eroavaa versiota. Seuraavaksi määritellään kuvan sisältämät termit.

Profiilin hyödyntäminen	
	Käyttäjien luokittelu ryhmiin
	Vertailu käyttäjien kesken (henkilölähtöinen)
	Vertailu käyttäjien kesken (objektilähtöinen)

Kuva 11. Profiilin hyödyntäminen

Käyttäjiä voidaan luokitella ryhmiin ominaisuuksiensa mukaan. Näitä ominaisuuksia voivat olla esimerkiksi sukupuoli tai auton vuosimalli. Erilaisia ryhmiä voidaan käyttää esimerkiksi kohdistamaan mainonta tietyntyyppisille asiakassegmenteille.

Käyttäjiä voidaan vertailla keskenään, jotta saataisiin selvitettyä käyttäjät, jotka ovat esimerkiksi kiinnostuneita samoista tuotteista. Tässä vertailussa on kaksi erilaista lähestymistapaa: henkilölähtöinen ja objektilähtöinen vertailu. *Henkilölähtöinen* tapa vertailee esimerkiksi käyttäjien tekemiä arviointeja keskenään toisiin käyttäjiin, jotka ovat arvostelleet objekteja samalla tavalla, kun taas *objektilähtöinen* tapa vertailee samalla tavalla arvosteltuja objekteja toisiinsa. Näistä objektikohtainen lähestymistapa on havaittu paremmaksi. Tämä lähestymistapa nostaa hieman arvioiden tarkkuutta verrattuna käyttäjäkohtaiseen lähestymistapaan. Erityisesti se kuitenkin lisää järjestelmän tehokkuutta, koska vaadittavat laskelmat voidaan suorittaa offline-tilassa (Sarwar, Karypis, Konstan ja Reidl 2001). Luvussa 5 esiteltävä MovieLens elokuvien suosittelujärjestelmä käyttää tällaista objektilähtöistä vertailua suosittaakseen käyttäjille sopivia elokuvia.

4.6 Profilointimenetelmässä käytettävät standardit ja suositukset

Profilointimenetelmän käyttämät standardit ja suositukset ovat avuksi niitä tutkittaessa. Standardit ovat tarkkaan määriteltyjä tapoja tehdä asioita, joten näiden käyttämisen pitäisi selkeyttää profilointimenetelmän toimintaa. Suositukset ovat käytännön standardeja (de facto) ja standardit ovat standardointijärjestöjen määrittämiä oikeita standardeja (de jure). Tässä työssä esiteltyihin menetelmiin on sisältynyt seuraavia standardeja ja suosituksia: P3P, XML (kuva 12). Seuraavaksi määritellään kuvan sisältämät termit.

Profilointimenetelmässä käytettävät standardit ja suositukset		
	P3P	
	XML	

Kuva 12. Profilointimenetelmässä käytettävät standardit ja suositukset

XML-kieli (Extensible Markup Language (XML) 1.0 (Fourth Edition) 2006) on yleisesti käytetty merkkauskieli. Kielellä kuvataan tietoa lisäämällä sen sekaan tietoa tarkentavia merkintöjä. Sen voikin kuvata sisältävän tietoa tiedosta. XML-kieltä pystyy lukemaan niin kone kuin ihminenkin. XML-kieltä on käytetty standardoimaan tietoa erilaisten tietoteknistenjärjestelmien integroinnissa, joten sen käyttäminen on eduksi myös profiloinnissa.

P3P (Platform for Privacy Preferences (P3P) Project 2006) on W3C:n suositus tietosuojakysymysten selkeyttämiseen käyttäjien ja kauppapaikkojen välillä. Suositusta tukevia kauppapaikkoja sekä selaimia (ainakin Internet Explorer 6 ja Mozilla Firefox) on useita (Teltzrow ja Kobsa 2004). Kauppapaikkojen kohdalla P3P:tä tukevat kaupat eivät ole enemmistönä. Egelmanin, Cranorin ja Chowdhuryn (2005) suorittaman tutkimuksen perusteella suunta näyttäisi kuitenkin olevan ylöspäin. P3P:a on kritisoitu muun muassa liiasta luottamisesta kauppapaikan tekijöihin tutkiessaan sen tietosuojaselostetta. Toisaalta tämä ongelma on aina kyseistä selostetta tutkittaessa, näin ollen se ei varsinaisesti rajoitu vain P3P:n luotettavuuteen (Wikipedia, 2006). P3P:n käytön lisäongelmana on se, että harvat käyttäjät tutkivat lainkaan kauppapaikkojen tietosuojaselosteita (Teltzrow ja Kobsa 2004). Standardin toimintaa kuvattiin tarkemmin kohdassa 2.3.3.

Luvussa muodostettiin viitekehys profilointimenetelmien ominaisuuksien vertailuun (kokonaisuudessaan liitteessä A). Tämän Laverin viitekehyksen tarkoituksena on toimia selkeänä kehikkona www-ympäristössä käytettävien profilointimenetelmien ominaisuuksien tutkimiseen. Viitekehystä voidaan käyttää myös hahmottamaan, minkälaisia ominaisuuksia erilaiset profilointimenetelmät voivat sisältää. Seuraavassa luvussa Laverin viitekehystä verrataan kahteen muuhun profilointimenetelmien tutkimisen viitekehykseen MovieLens-suosittelujärjestelmän avulla.

5 VIITEKEHYKSEN TESTAUS

Tämä luku keskittyy luvussa neljä muodostetun viitekehyksen testaamiseen. Viitekehystä verrataan Middletonin (2003, 20) sekä Wun, Tremainen, Instonen ja Turoffin (2002) esittämiin viitekehyksiin. Molemmat näistä viitekehyksistä ovat suunnattuja profilointimenetelmien ominaisuuksien tunnistamiseen. Luvussa tutkitaan jokaisen kolmen viitekehyksen näkökulmasta samaa profilointijärjestelmää. Tämä järjestelmä on avoimeen lähdekoodiin perustuva MovieLens-elokuvien suosittelujärjestelmä. Luku vastaa Nunamakerin, Chenin ja Purdinin (1990) esittämässä mallissa (katso 1.3) kohtaa testaus. Testauksessa pyritään tutkimaan edellisten vaiheiden tuloksia ja mahdollisesti löytämään niistä ongelmia, joita voidaan myöhemmin ratkaista (Nunamaker ym 1990, 95).

Tässä työssä ei tarkoituksellisesti valittu tutkittavaksi mitään profilointia hyödyntävää kauppapaikkaa. Syy tähän on yksinkertaisesti se, että näistä kaupallisista järjestelmistä ei saa tietoa joko lainkaan tai vähintäänkin huonosti. Tämän lisäksi luvun tarkoituksena ei ole varsinaisesti testata järjestelmää vaan vertailla kolmea erilaista viitekehystä toisiinsa. Juuri tästä syystä MovieLens on hyvä valinta: johtuen sen akateemisesta käyttötarkoituksesta ja taustasta, sen kehitys ja ominaisuudet ovat hyvin ja avoimesti dokumentoituja.

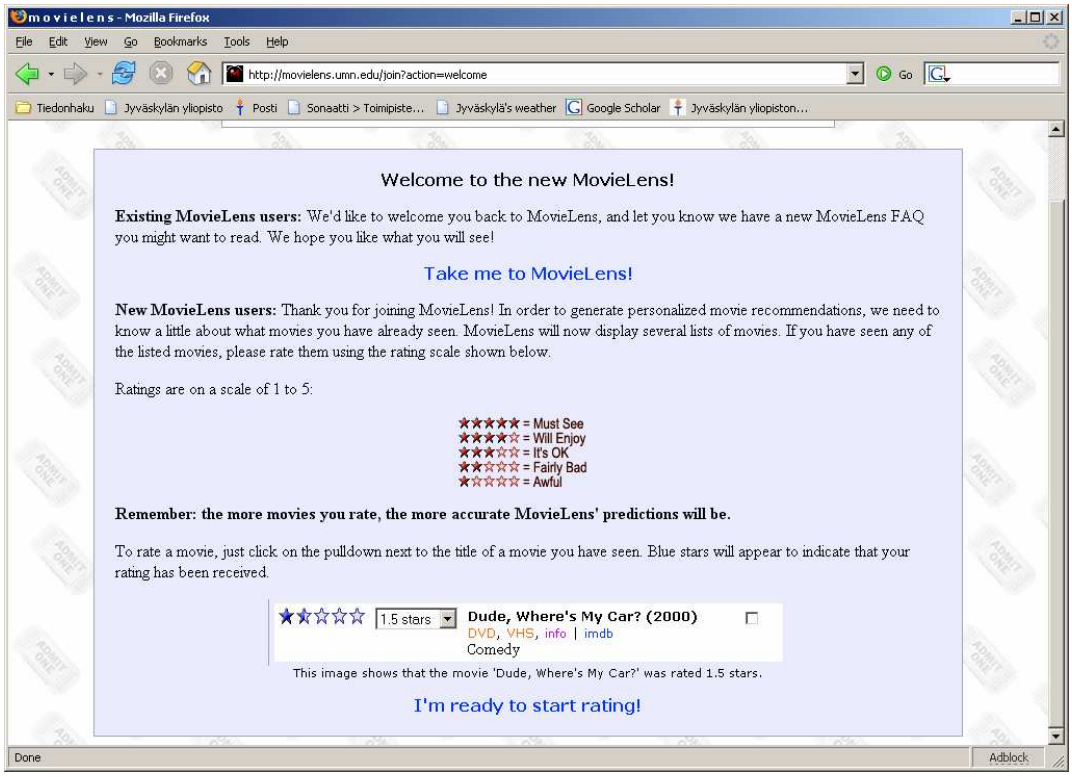
Seuraavaksi esitellään ensin MovieLens-järjestelmä ja tämän jälkeen käydään jokainen viitekehys läpi yksi kerrallaan. Jokaisen viitekehyksen kohdalla esitellään ensin itse viitekehys, jonka jälkeen tutkitaan MovieLens-järjestelmää kyseisen kehyksen avulla. Kun kaikki viitekehykset on käyty läpi, tehdään vielä yhteenveto näiden kolmen eri viitekehyksen eroista.

5.1 MovieLens-suosittelujärjestelmä

Kolmea viitekehystä vertaillaan toisiinsa tutkimalla niistä jokaisella erikseen MovieLens-suosittelujärjestelmää. MovieLens on osana GroupLens-projektia Minnesotan yliopistossa tehty elokuvien suosittelujärjestelmä (Rashid, Albert,

Cosley, Lam, McNee, Konstan ja Riedl 2002). MovieLens on avoimeen lähdekoodiin perustuva projekti, jonka tarkoituksena on tutkia ja testata GroupLensissä kehitettyä teknologiaa, sekä yleisesti profilointiin liittyviä aiheita.

Kuvassa 13 näkyy MovieLens-järjestelmän esittely. MovieLens tarjoaa käyttäjille mahdollisuuden arvostella elokuvia antamalla niille yhdestä viiteen tähteä. Näiden tähtien avulla järjestelmä myös kertoo arvionsa siitä, kuinka paljon käyttäjä pitää elokuvasta, jota hän ei ole vielä nähnyt tai arvostellut.



Kuva 13. MovieLens-suosittelujärjestelmä.(MovieLens)

MovieLens tunnistaa käyttäjänsä rekisteröitymisen ja sisäänkirjautumisen avulla. Evästettä voidaan myös käyttää muistamaan käyttäjä rekisteröitymisen jälkeen käyttäjän itsensä niin halutessa. Muut käytettävät evästeet ovat vain teknisiä apuvälineitä, esimerkiksi istunnon tietojen välittämistä järjestelmän eri sivu-

jen kesken. Käyttäjät voivat saada suosituksia myös ilman rekisteröitymistä palveluun, mutta rekisteröitymällä he saavat tarkempia suosituksia ja käyttöönsä erilaisia profiloinnin työkaluja. GroupLensin kotisivuilla on saatavilla vapaasti kaksi järjestelmästä kerättyä erikokoista kantaa elokuvien arvosteluista (GroupLens 2003).

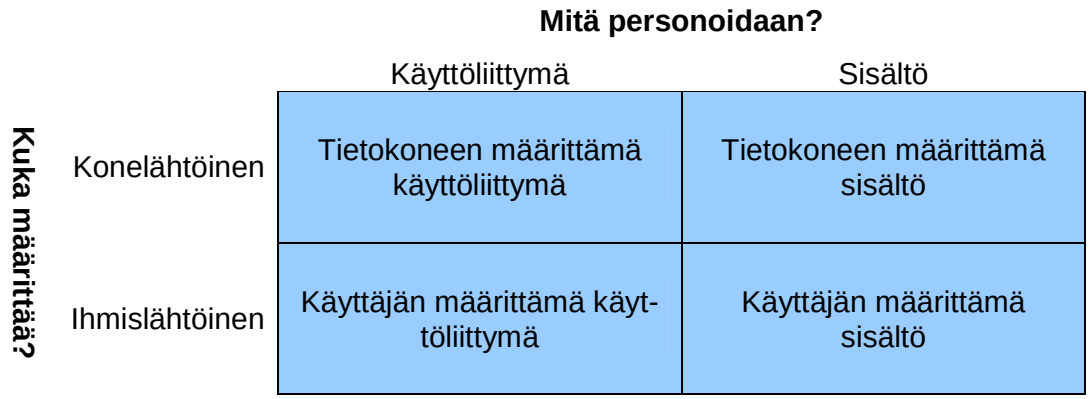
MovieLens on paisunut varsin mittavaksi huomioon ottaen sen, että se on kehitetty pelkästään tutkimusta varten. Yksittäisiä kävijöitä sivustolla oli esimerkiksi vuoden 2006 alussa kuukauden aikana lähes 3400 (Sen, Lam, Rashid, Cosley, Frankowsky, Osterhouse, Harper ja Riedl 2006). Tämä ei ole tietenkään paljon esimerkiksi verrattuna Amazon.comiin, mutta aivan riittävä määrä tämäntyyppiselle sivustolle ja tätä tutkimusta varten.

5.2 MovieLens Wun, Tremainen, Instonen ja Turoffin viitekehyksen mukaan

Wun ym. (2002) esittelemä viitekehys on kolmesta tässä luvussa käytetystä viitekehyksestä yksinkertaisin. Suuri osa profilointimenetelmien luokittelun viitekehyksistä on tällaisia vastaavanlaisia nelikenttiä. Näissä nelikentissä eivät käytetyt termit välttämättä ole samoja, mutta tarkkuus lopputuloksen kanssa on vastaava kuin Wun ym. (2002) viitekehyksessä. Esimerkkinä toisesta tällaisesta nelikentästä on luvussa kolme esitelty Eirinakin ja Vazirgianniksen (2003) esittämä profilointimenetelmien jako. Eirinakin ja Vazirgianniksen mallissa keskitytään enemmän siihen, miten profilointi suoritetaan, kun taas Wu ym. (2002) keskittyvät enemmän tämän profiloinnin yhteen osaan, personointiin. Sinänsä tämä on toimiva lähtökohta, koska huomattava osa profiloinnin hyödyntämistavoista on personointia.

Wun ym. (2002) viitekehys esitetään kuvassa 14. Kuvassa näkyy, kuinka viitekehys tutkii järjestelmiä kahdesta eri näkökulmasta: mitä personoidaan ja kuka määrittää personoitavan tiedon. Nämä kaksi jakautuvat tämän jälkeen osiin seuraavalla tavalla. ”Mitä personoidaan” koostuu käyttöliittymän tai sisällön

personoinnista. ”Kuka määrittää” koostuu konelähtöisestä ja ihmislähtöisestä personoinnista. Seuraavaksi kuvataan viitekehyksen ominaisuudet.



Kuva 14. Wu ym. (2002) viitekehys profilointimenetelmien tunnistamiseksi

Käyttöliittymän personointi on näkyvän tiedon uudelleenjärjestelyä käyttäjälle yksilöidyllä tavalla (Wu ym. 2002). Käyttöliittymän personoinnin alle ei kuulu tiedon muokkaus vaan ainoastaan ulkoasun muuttaminen esimerkiksi käyttäjän profiilin mukaiseksi.

Sisällön personointi on sen sijaan tiedon tai tietolinkkien sisällön uudelleenmuokkausta käyttäjän profiilin mukaan (Wu ym. 2002). Tällöin tiedon ulkonäön lisäksi, myös sen sisällön on muututtava.

Konelähtöisessä profiloinnissa aktiivinen personoinnin kehittäjä on tietokone (Wu ym. 2002). Tällöin kone kerää tietoja seuraamalla esimerkiksi käyttäjän toimia sivulla, ostettuja tuotteita tai jollakin kyseisellä kauppapaikan sivulla vietettyä aikaa ilman, että käyttäjä on täysin tietoinen tästä prosessista.

Ihmislähtöisessä profiloinnissa on käyttäjä aktiivisesti ja tietoisesti mukana kehittämässä profiloinnin lopputulosta (Wu ym 2002). Tällainen osallistuminen voi olla esimerkiksi eri objektien arvostelua tai vaikkapa kiinnostuskohdelistojen täyttämistä.

Wu ym. esittelivät viitekehyksen yhteydessä erilaisia profiloinnin hyödyntämisominaisuuksia, joita voidaan käyttää profilointiasteen määrittämisessä. Nämä ominaisuudet ovat personoinnin hallinta, sisällön personointi, linkkien personointi, asiakaskohtainen ulkoasun muokkaaminen sekä ihmismäinen käyttöliittymä. Näitä ominaisuuksia on esitelty tarkemmin tämän työn kohdassa 2.2.2 samassa yhteydessä muiden profiilin hyödyntämistapojen kanssa.

Wu ym. (2002) käyttivät viitekehystään tutkimaan erilaisten sivujen ja järjestelmien profiilin hyödyntämistapoja. Tämän jälkeen erilaiset tavat pisteytettiin niiden profiloinnin tason mukaan kuvassa 15 esitellyn pisteyttämispohjan avulla. Mitä korkeammat pisteet, sitä korkeampi on personoinnin taso. Esimerkiksi ”ihmislähtöisen personoinnin hallinnan” (kolme pistettä) katsotaan nostavan yleistä profiloinnin tasoa enemmän kuin ”ihmismäisen käyttöliittymän” (yksi piste). Kun kaikki ominaisuudet on käyty läpi, lasketaan pisteet yhteen. Tulokseksi saadaan yhteispistemäärä, jonka perusteella voidaan päätellä kuinka iso on järjestelmän tai kauppapaikan profiloinnin aste. Tällä tavalla pisteytetyt järjestelmät eivät luonnollisestakaan ole verrannollisia muihin kuin samalla pisteytyspohjalla pisteytettyihin järjestelmiin.

Konelähtöinen personointi	
Ominaisuus	Pisteet
Sisällön personointi hyödyntäen konelähtöistä tietoa	3
Konelähtöinen profiloinnin hallinta	3
Automaattinen linkkien personointi	2
Ihmismäinen käyttöliittymä	1
Ihmislähtöinen personointi	
Ominaisuus	Pisteet
Sisällön personointi hyödyntäen käyttäjän arviointeja	3
Ihmislähtöinen personoinnin hallinta	3
Käyttäjän muokkaama linkkien personointi	2
Asiakaskohtainen ulkoasun muokkaaminen	1

Kuva 15. Pisteytyspohja Wu ym (2003) viitekehykselle

Seuraavaksi tällä samalla periaatteella tutkitaan MovieLens-suosittelujärjestelmää. Ensin tarkastellaan järjestelmää sen perusteella, mitä profiloidaan. Tämän jälkeen tutkitaan näitä havaittuja ominaisuuksia sen pohjalta, miten profilointi suoritetaan. Erilaisille ominaisuuksille annettavat pisteet ovat suoraan kuvan 15 taulukosta.

Vertailun tulokset löytyvät taulukkona tutkielman liitteestä B.

5.2.1 Mitä profiloidaan

Tässä kohdassa tutkitaan, mitä profiloidaan. Tällöin tunnistetaan ominaisuuksia, joita järjestelmä muokkaa jokaista käyttäjää varten erikseen tämän profiilia vastaavalla tavalla. Nämä tunnistettavat ominaisuudet ovat edellä mainittuja profiilin hyödyntämistapoja.

Kuten aikaisemmin todettiin, MovieLens tarjoaa suosituksia käyttäjien arvosteluiden pohjalta vertailemalla samalla tavalla arvosteltuja elokuvia keskenään. Tätä voidaan edellä mainittujen ominaisuuksien mukaan kutsua *sisällön personoinniksi*. Pisteitä tämä saa kuvan 15 pisteytyspohjan mukaan kolme.

Kirjauduttaessa sisään järjestelmään käyttäjää tervehditään nimellä sekä annetaan muita käyttäjään liittyviä tietoja, kuten arvosteltujen elokuvien määrä. Tämä toimintatapa luetaan hyödyntämistavan *ihmismäinen käyttöliittymä* alle. Tämä ominaisuus saa kuvan 15 pisteytyspohjan mukaan yhden pisteen.

Järjestelmä ei anna käyttäjän muokata käyttöliittymän ulkoasua, joten se ei hyödynnä *asiakaskohtaista ulkoasun muokkausta*. Sivusto ei myöskään sisällä *linkkien personointia*. Linkkien personointi tarkoittaa lähinnä turhien linkkien poistoa tai hyödyllisten linkkien lisäämistä näkyviin sivustolle. Muutokset, jotka näkymässä tapahtuvat, ovat luokiteltavissa lähinnä sisällönmuutoksiksi.

5.2.2 Kuva määrittää

Edellisessä kohdassa tunnistettiin ominaisuuksia, jotka sivustolla toteuttavat profilointia. Seuraavaksi tutkitaan jokaisesta ominaisuudesta, miten se on muodostettu. Tarkemmin sanoen onko sen muodostanut käyttäjätietoisesti vai onko se ennemminkin järjestelmän luomaa. Saattaa olla ominaisuuksia, jotka eivät kuulu selkeästi kumpaankaan mainituista ryhmistä, sisältäen piirteitä näistä molemmista.

MovieLens-järjestelmän todettiin sisältävän seuraavat ominaisuudet: ihmismäinen käyttöliittymä ja sisällön personointi. Ihmismäinen käyttöliittymä on selkeästi tietokoneen luomaa eli konelähtöistä. Sisällön personointi saattaa sen sijaan saattaa olla joskus vaikeastikin luokiteltavaa. MovieLensin tapauksessa se on käyttäjän aktiivisella osallistumisella luotua, ollen siis ihmislähtöistä. Järjestelmä kuitenkin suosittaa elokuvia automaattisesti myös esimerkiksi käyttäjän kirjautuessa sisään, joten myös *konelähtöisen sisällön personoinnin* tunnusmerkit täyttyvät. Näin ollen järjestelmä saa kolme lisäpistettä kuvan 15 pisteytyspohjan mukaan.

Viitekehyksen pisteytyksen mukaan MovieLens-järjestelmä sai yhteensä seitsemän pistettä. Kuten aikaisemmin todettiin, ei tämä kerro juuri mitään, ennen kuin tulosta verrataan johonkin toiseen samaan viitekehykseen suhteutettuun järjestelmään. Esimerkkinä kaupallisista järjestelmistä Amazon.com sai samalla viitekehyksellä tutkittuna 14 pistettä (Wu ym 2002).

5.3 MovieLens Middletonin viitekehyksen mukaan

Middletonin (2003, 20) esittämä profilointimenetelmien tunnistamisen viitekehys on jo huomattavasti tarkemmin menetelmien yksittäisiä ominaisuuksia tarkasteleva kuin edellä esitetty Wun ym. (2002) kehittämä versio. Middletonin kehys on joiltakin osin vastaava tässä työssä kehitetyn viitekehyksen kanssa.

Middletonin kehys on kuitenkin joiltain osin teoreettisempi, keskittyen esimerkiksi tulosten takana oleviin erilaisiin malleihin tarkemmin.

Middletonin (2003) viitekehys esitellään kuvassa 16. Kuvassa näkyy kuinka viitekehys koostuu viidestä eri pääkohdasta, joiden avulla pyritään selvittämään profilointimenetelmän toimintaa. Nämä viisi kohtaa ovat profilointimenetelmä, jaettu tieto, profiilin esitys, tietolähde ja suositteluteknikka. Näistä jokainen jakaantuu useampaan alakohtaan. Seuraavaksi esitellään näiden kohtien sisällöt ja merkitykset lyhyesti. Seuraavaksi kuvataan viitekehyksen ominaisuudet.

Profilointimenetelmä			
	Konelähtöinen		
		- Toiminnan seuraaminen - Tietoa päättelevät heuristiikat	
	Ihmislähtöinen		
		- Käyttäjien palaute - Käyttäjien suorittama ohjelmointi	
			- Suodatussäännöt - Käyttäjien luomat ryhmät/kategoriat
Jaettu tieto			
	Käyttäjien palaute		
		- Palaute objekteista - Esimerkit objekteista - Selaushistoria	
	Järjestelmätieto		
		- Objekti ryhmät/kategoriat - Järjestelmäheuristiikat	
Profiilin esitys			
	Vektorimalli Selausjäljet Tietämyspohjainen profiili		
Tietolähde			
	Sisäinen objektitietokanta Tutkitut www-sivut Järjestelmän ulkopuoliset tapahtumat		
Suositteluteknikka			
	Heuristiikat Samanlaisuuksien yhdistäminen Yhteistoimintapohjainen suodattaminen		

Kuva 16. Middletonin (2003, 19) viitekehys profilointimenetelmien vertailuun

Profilointimenetelmä koostuu konelähtöisestä ja ihmislähtöisestä profiloinnista. Konelähtöinen profilointi jakaantuu alakohtiin toiminnan seuraaminen sekä tietoa päättelevät heuristiikat. Ihmislähtöinen profilointi jakaantuu käyttäjien palautteeseen ja käyttäjien suorittamaan ohjelmointiin. Käyttäjien suorittama ohjelmointi jakaantuu vielä suodatussääntöihin ja käyttäjien luomiin ryhmiin/kategorioihin. *Toiminnan seuraaminen* (engl. monitoring behaviour) tarkoittaa yksinkertaisesti sitä, että järjestelmä seuraa ja tallentaa sen, miten käyttäjät toimivat järjestelmän kanssa. *Tietoa päättelevät heuristiikat* (engl. heuristics to infer information) ovat sääntöjä, joiden avulla käyttäjistä päätellään tietoa. *Käyttäjien palaute* (engl. user feedback) on käyttäjien antamaa yksiselittäistä palautetta objekteista, koskien esimerkiksi jonkun kyseisen objektin hyödyllisyydestä itselleen tai kertomalla esimerkkejä heitä kiinnostavista objekteista. *Suodatussäännöt* (engl. filter rules) ovat käyttäjien määrittämiä objektien suodattamissääntöjä. *Käyttäjien luomat ryhmät/kategoriat* (engl. user-created groups/categories) ovat nimensä mukaisesti objektien jaottelua käyttäjän itsensä määrittämällä tavalla. (Middleton 2003, 20)

Jaettu tieto koostuu kohdista käyttäjien palaute ja järjestelmätieto. Käyttäjien palaute jakaantuu alakohtiin palaute objekteista ja esimerkit objekteista. Järjestelmätieto jakaantuu alakohtiin objekti ryhmät/kategoriat ja järjestelmäheuristiikka. *Palaute objekteista* (engl. item feedback) on käyttäjien avuksi tarkoitettua järjestelmän antamaa palautetta. *Esimerkit objekteista* (engl. examples of items) ovat järjestelmän muodostamia harjoitusryhmiä objekteista, joita käytetään ymmärtämään tarkemmin käyttäjien toiveet. *Objekti ryhmät/kategoriat* (engl. item groups/categorizations) ovat järjestelmän jakamia yhteisiä objektijoukkoja, jotka voivat olla joko koneellisesti tai käyttäjien toimesta määritettyjä. *Järjestelmäheuristiikka* (engl. domain heuristics) ovat suodatussääntöjä, jotka järjestelmä jakaa kaikkien käyttäjien kesken. (Middleton 2003, 20)

Profiilin esitys koostuu seuraavista alakohdista: vektorimalli, selausjäljet ja tietämyspohjainen profiili. *Vektorimallissa* (engl. vector model) järjestelmä käyttää vektoreita mallintaakseen dokumentteja tai profiileita. *Selausjäljet* (engl. navigation trails) ovat käyttäjän jättämiä navigointijälkiä, joita järjestelmä säilyttää. *Tietämyspohjainen profiili* (knowledge-based profile) on profiili, joka sisältää vahvistettuja faktoja käyttäjästä ja jonka pohjalta voidaan vetää johtopäätöksiä käyttäjäryhmistä ja näiden kiinnostusalueista. (Middleton 2003, 20)

Tietolähde koostuu seuraavista alakohdista: sisäinen tietokanta objekteille, tutkitut www-sivut ja järjestelmän ulkopuoliset tapahtumat. *Sisäinen objektitietokanta* (engl. internal database of items) on järjestelmän sisäinen tietokanta, jota hyväksi käyttäen objekteja suositellaan käyttäjille. *Tutkitut www-sivut* (engl. crawled web pages) ovat järjestelmän läpikäymiä sivustoja, joilta voidaan suositella objekteja. *Järjestelmän ulkopuoliset tapahtumat* (engl. external domain events) ovat järjestelmään suoraan kuulumattomia tapahtumia, kuten sähköpostin saapuminen, jotka saavat aikaan objektien suosittamisen. (Middleton 2003, 20)

Suositteluteknikka jakaantuu alakohtiin heuristiikat, samankaltaisuuksien yhdistäminen ja yhteistoimintapohjainen suodattaminen. *Heuristiikat* (engl. heuristics) tarkoittavat, että käytetään sääntöjä, joiden avulla pyritään löytämään parhaat suositukset kyseiselle käyttäjälle. *Samankaltaisuuksien yhdistämisessä* (engl. similarity matching) käytetään *samankaltaisuus funktiota* (engl. similarity function) löytämään sisältöpohjaisen profiilin mukaisia objekteja. *Yhteistoimintapohjaisessa suodattamisessa* (engl. collaborative filtering) käytetään tilastollisia menetelmiä valitsemaan suositeltavat objektit samankaltaisista objekteista pitävien käyttäjien avulla. Toisin kuin esimerkiksi kohdassa 3.1.1 esitelty yhteistoimintapohjainen profilointi, ei Middleton ota yhteistoimintapohjaisessa suodattamisessa kantaa siihen, miten profiili on muodostettu. (Middleton 2003, 20)

Seuraavaksi suoritettavan vertailun tulokset löytyvät taulukkona liitteestä C.

5.3.1 Profilointimenetelmä

Profilointimenetelmän alla ovat profiilinmuodostamisvaiheessa käytettävät ominaisuudet tai menetelmät, joiden pääasiallinen tehtävä on kerätä tietoa profiilia varten. Muut ominaisuudet sijoittuvat kahden, edellä jo useaan kertaan mainitut pääkohdan alle. Nämä pääkohdat ovat konelähtöinen ja ihmislähtöinen profilointi. (Middleton 2003, 17)

Profilointimenetelmänä Movielens-järjestelmässä käytetään *ihmislähtöistä menetelmää* eli käyttäjä on aktiivisena ja tietoisena osana profiilin muodostamista. Ihmislähtöisen menetelmän ominaisuuksista käytetään erityisesti *käyttäjäpalautetta* avuksi. Ominaisuutta hyödynnetään, kun käyttäjät arvostelevat jo näkemiään elokuvia. Muita profilointimenetelmiä ei käytetä.

5.3.2 Jaettu tieto

Kohdassa ”jaettu tieto” tutkitaan, mitä tietoa järjestelmä jakaa muodostaakseen profiilia tai hyödyntääkseen sen sisältämiä tietoja paremmin. Jaettu tieto jakaantuu joko käyttäjien palautteeseen tai järjestelmän tietoon. (Middleton 2003, 17-18)

Kohdan ”käyttäjien palaute” sisältämistä ominaisuuksista, MovieLens-järjestelmä käyttää *palautetta objekteista*. MovieLens-käyttäjän suositusten pohjana hyödynnetään jo olemassa olevien käyttäjien aikaisemmin arvostelemia objekteja.

Kohdan ”järjestelmätieto” sisältämistä ominaisuuksista, MovieLens-järjestelmä käyttää *objektien ryhmiä/kategorioita*. Järjestelmä jakaa elokuvia niiden ominaisuuksien mukaan joukkoihin useilla eri tavoilla. Näitä ominaisuuksia käytetään hyödyksi yritettäessä laskea suosituksia käyttäjille.

5.3.3 Profiilin esitys

Kohdassa ”profiilin esitys” tutkitaan, miten profiili on järjestelmään mallinnettu. Tällä mallinnustavalla saattaa olla suurtakin merkitystä muun muassa profiilin tehokkuuden kannalta. Mallintamistapoja on kolme erilaista: vektorimalli, navigointijäljet sekä tietämyspohjainen profiili. Kaikki nämä tavat voivat olla käytössä yhtä aikaa eivätkä sulje toisiaan pois. (Middleton 2003, 18)

MovieLens-järjestelmä käyttää profiilin esittämiseen *vektorimallia* (Sobecki 2005). Vektorimalli on normaali ja yleisesti käytössä oleva tapa esittää profiili, koska siihen pystytään helposti hyödyntämään koneoppimisen keinoja (Middleton 2003, 18). Järjestelmä ei hyödynnä selausjälkiä tai tietämyspohjaista profiilia.

5.3.4 Tietolähde

Kohdassa ”tietolähde” tutkitaan, minkälaisia tietolähteitä järjestelmä käyttää muodostettaessa tai hyödynnettäessä profiilia. Hyödyntämistapoja on kolme: sisäinen objektitietokanta, tutkitut www-sivut sekä järjestelmän ulkopuoliset tapahtumat. (Middleton 2003, 18)

MovieLens-järjestelmä hyödyntää tietolähteenään *sisäistä objektitietokantaa*. Tässä tietokannassa säilytetään tiedot käyttäjistä, elokuvista ja käyttäjien tekemistä arvosteluista.

5.3.5 Suositteluteknikka

Kohdassa ”suositteluteknikka” tutkitaan sitä, kuinka käyttäjille tarjottavat suositukset muodostetaan. Suurin osa käytössä olevista tekniikoista voidaan jakaa kolmeen eri luokkaan (Middleton 2003, 18). Nämä luokat ovat heuristiikoiden käyttäminen, samanlaisuuksien yhdistäminen sekä yhteistoimintapohjainen suodattaminen.

MovieLens-järjestelmä käyttää suositusten luomiseen *yhteistoimintapohjaista suodattamista*. Käyttäjien arvostellessa elokuvia yritetään löytää käyttäjiä, jotka ovat pitäneet samanlaisista elokuvista. Tämän tuloksen perusteella suositellaan käyttäjälle uusia elokuvia sen perusteella, miten muut käyttäjät ovat näkemiään elokuvia arvostelleet.

5.4 MovieLens Laverin viitekehyksen mukaan

Luvussa neljä kehitetty viitekehys löytyy kokonaisuudessaan liitteestä A. Jokaisen arviointikohdan tarkemmat selitykset löytyvät luvusta 4. Viitekehyksen tarkoituksena on tarkastella profilointimenetelmien varsinaisia ominaisuuksia varsin yksityiskohtaisella tasolla, menemättä kuitenkaan liian syvälle menetelmän tekniseen puoleen.

MovieLens-järjestelmää tarkastellaan seuraavaksi tässä työssä kehitetyn viitekehyksen sisältämän kuuden kohdan avulla. Nämä kohdat ovat muodostusvaihe, ylläpitovaihe, tiedonsuodatusmenetelmä, kerättyjen tietojen hyödyntäminen käyttäjän tunnistus sekä käytetyt standardit.

Seuraavaksi suoritettavan vertailun tulokset löytyvät taulukkona liitteestä D.

5.4.1 Käyttäjän tunnistus

Kohdassa ”käyttäjän tunnistus” tutkitaan, miten profilointijärjestelmä erottaa käyttäjät toisistaan. Luotettava tunnistus on olennaisen tärkeää oikean profiilin luomiselle. Tunnistuksen toimivuuteen vaikuttavia tekijöitä ovat käyttäjän käyttämät erilaiset päätelaitteet tai eri paikoista suoritetut kauppapaikan käytöt. Samalta koneelta samaa kauppapaikkaa saattaa käyttää myös useampi eri käyttäjä. Tässä työssä kuvatut käyttäjän tunnistusmenetelmiä ovat evästeet, sisäänkirjautuminen sekä kolmannen osapuolen hallussa oleva profiili.

MovieLens-järjestelmässä käyttäjät tunnistetaan rekisteröitymällä ja myöhemmin *sisäänkirjautumalla sivustolle*. Järjestelmää voidaan käyttää osittain myös ilman rekisteröitymistä, mutta tämä ei ole suositeltavaa tarkkuuden kannalta. Ilman rekisteröitymistä käyttäjän tekemät toimet eivät jää mitenkään talteen. Sisäänkirjautuminen on näistä kolmesta tunnistustavasta luotettavin. Ainoa huono ominaisuus tässä tavassa on, että käyttäjälle tulee hieman lisää työtä.

5.4.2 Muodostusvaihe

Muodostusvaiheessa tutkitaan ominaisuuksia, joiden avulla profiilin ensimmäiset tiedot luodaan. Tämä vaihe on selkeästi kriittisin profiilin elämänkaaressa. Profiloinnilla on tässä vaiheessa suurimmat mahdollisuudet mennä pieleen esimerkiksi puutteellisten tai virheellisten tietojen takia. Tai mahdollisesti jopa siksi, että mitään tietoja ei vielä ole olemassa. Tämän vaiheen ominaisuuksia ovat mahdollisuus tuoda valmiita tietokantoja, ihmislähtöinen ja konelähtöinen profilointi. Sekä ihmislähtöinen että konelähtöinen profilointi jakaantuvat molemmat yksityiskohtaisempiin ominaisuuksiin, jotka on kuvattu tarkemmin luvussa neljä.

MovieLensia perustettaessa käytettiin HP:n ja Compaqin lopetetun tutkimuskohteen EachMovie-suositusjärjestelmän jälkeensä jättämää tietokantaa (GroupLens 2003). Järjestelmään on siis *mahdollista tuoda olemassa olevia tietokantoja*. MovieLens-järjestelmän käyttäjillä on mahdollisuus arvostella elokuvia. Tähän järjestelmä käytännössä perustuukin. Käyttäjät arvostelevat elokuvia, jonka jälkeen järjestelmä vertailee näitä arvosteluita keskenään ja näiden tulosten perusteella suosittaa käyttäjille elokuvia, joista he mahdollisesti pitävät. Käyttäjien täytyy arvostella tietty määrä elokuvia ennen kuin MovieLens ryhtyy tarjoamaan suosituksia. Järjestelmä suosittaakin, että käyttäjät arvostelisivat vähintään 35 elokuvaa ennen kuin tarkastelevat omia elokuvasuosituksiaan. Suositteluvaihe kuuluu profiilin hyödyntämiseen ja käsitellään siellä uudestaan.

5.4.3 Ylläpitovaihe

Ylläpitovaiheessa profilointijärjestelmän tärkeimmät tehtävät ovat profiilin tarkentaminen sekä käyttäjän muuttuvien tarpeiden mukana pysyminen. Profilointijärjestelmissä voidaan käyttää vain yhtä menetelmää koko profiilin elinkaaren ajan, jolloin ylläpitovaihe ei eroa muodostusvaiheesta lainkaan.

MovieLens-järjestelmässä ylläpitovaihe ei eroa muodostusvaiheen profiloinnista. Profiilien ylläpitoon käytetään samaa menetelmää kuin muodostusvaiheessakin. Käyttäjien täytyy tietysti myös arvostella elokuvia aktiivisesti, jotta järjestelmä pysyisi käyttäjien makujen mukana.

5.4.4 Tiedonsuodatusmenetelmä

Kohdassa ”tiedonsuodatusmenetelmä” tutkitaan, minkälaisia menetelmiä profilointijärjestelmä käyttää suodattaakseen tietoa sen keräämästä käsittelemättömästä tiedosta. Nämä menetelmät saattavat olla matemaattisia algoritmeja tai kyselykieliä. Algoritmit voivat joissakin tapauksissa olla profilointimenetelmän perusta (mm. Eirinaki ja Vazirgiannis 2003, sekä Yu, Tresp ja Yu 2004). Tässä työssä tunnistettuja tiedonsuodatusmenetelmiä ovat XML-QL, Bayesin teoreeman johdannaiset sekä olettamusfunktion johdannaiset.

MovieLens käyttää suositusten luomiseen algoritmia nimeltään yhteistoimintapohjainen suodattaminen. Tämä algoritmi on kuitenkin osa yhteistoimintapohjaista profiilin muodostusta, ja näin ollen liittyy ennemminkin kohtiin muodostus- ja ylläpitovaihe.

5.4.5 Kerättyjen tietojen hyödyntäminen

Profiiliin kerättyjä tietoja olisi järkevä myös hyödyntää jollakin tavalla. Tässä vaiheessa tutkitaan, minkälaisin tavoin järjestelmä tätä hyödyntämistä toteuttaa. Tässä työssä kuvatut tiedonhyödyntämismenetelmät ovat käyttäjien luokit-

telu ryhmiin, käyttäjien vertailu (henkilölähtöinen) ja käyttäjien vertailu (objektilähtöinen).

MovieLens-järjestelmä käyttää edellä mainituista menetelmistä *objektilähtöistä käyttäjien kesken vertailua*. Tällaisessa tilanteessa käyttäjille tehdään suosituksia vertailemalla arvosteltuja objekteja ja etsimällä näistä yhtäläisyyksiä.

5.4.6 Käytetyt standardit ja suositukset

Kohdassa ”käytetyt standardit ja suositukset” tutkitaan, mitä standardeja profilointimenetelmä sisältää tai käyttää. Standardit auttavat profilointijärjestelmän integrointia muihin mahdollisesti olemassa oleviin järjestelmiin. Standardit helpottavat myös yleisesti profilointijärjestelmän toiminnan ymmärtämistä. Tässä työssä kuvattuja standardeja ja suosituksia olivat XML (Extensible Markup Language (XML) 1.0 (Fourth Edition) 2006) ja P3P (Platform for Privacy Preferences (P3P) Project 2006).

MovieLens järjestelmää pohjautuu osittain *XML*:n ja tämän liitännäiskieleen XSLT:n (XSL Transformations (XSLT) Version 1.0 2006). Järjestelmässä käytössä olevalla Java servlet-tasolla hyödynnetään myös XML-kieltä. (MovieLens)

5.5 Menetelmien erot

Seuraavaksi käsitellään kolmen tässä luvussa käsitellyn viitekehyksen eroja ja ominaisuuksia. Kaikkien viitekehysten yhteydessä kerrotaan, mihin ne soveltuvat ja mihin eivät. Tulokset näkyvät tiivistettyinä liitteessä E.

Wu ym. (2002) viitekehyksen avulla ei saada profilointijärjestelmästä todella yksityiskohtaista kuvaa. Lukuun ottamatta muutamia personoitavia ominaisuuksia saadaan tulokseksi lähinnä onko profilointi enemmän käyttäjä- vai konelähtöistä. Kuten aikaisemmin jo todettiin, viitekehys keskittyykin tutkimaan ennemmin profiloinnin yksittäistä osaa, personointia. Kuten odotettua, Wun ym. vii-

tekehys on sopiva yleisen personoinnin tason määrittämiseen, mutta ei niinkään yksityiskohtaiseen profilointimenetelmien tutkimiseen. Pisteytysjärjestelmän toiminta on jokseenkin sekava, johtuen ominaisuuksien selkeän määrityksen vaikeudesta. Tämän lisäksi näitäkin tuloksia voidaan verrata suoraan vain muihin saman viitekehyksen avulla vertailtuihin järjestelmiin.

Middletonin (2003) viitekehys on huomattavasti tarkempi kuin Wu:n ym. (2002) esittämä. Tässä yksityiskohtaisuudessa saattaa piillä myös ongelma. Tutkijan tarvitsee tutustua viitekehykseen varsin yksityiskohtaisesti ymmärtääkseen eri ominaisuuksien merkityksen. Kunnollisen tutustumisen jälkeen ei ongelmia ollut yhdessäkään kohdassa tutkittaessa kehyksen avulla MovieLens-järjestelmää. Viitekehys on kehitetty melko matemaattisesta näkökulmasta ja tämä näkyy useissa kehyksen sisältämissä ominaisuuksissa. Useat ominaisuuksista liittyvät erilaisiin matemaattisiin malleihin Vaikka viitekehys siihen kohtuullisen hyvin soveltuukin, ei sekään kuitenkaan tarkastele profilointimenetelmiä yleisesti. Lähtökohtaisesti se onkin kehitetty tutkimaan ainoastaan suosittelujärjestelmien ominaisuuksia.

Laverin viitekehyksen tarkoituksena on olla selkeä tapa tutkia ja ymmärtää profilointimenetelmiä. Viitekehys pyrkii olemaan ymmärrettävä ilman, että sen käyttäjän tarvitsee uppoutua profilointiin liian tarkasti. Kahdesta edellä mainitusta viitekehyksestä se muistuttaa lähemmin Middletonin esittämää versiota. Laverin viitekehys on näistä kahdesta kuitenkin monipuolisempi ja laajempi. Pyrkimyksestä huolimatta sillä on jossain määrin yhtenevä ongelma Middletonin kehyksen kanssa: yksityiskohtaisuudesta seuraa se, että viitekehykseen täytyy tutustua varsin tarkasti, jotta ymmärtää mitä mikäkin kohta tarkoittaa. Erona Middletonin viitekehykseen on tämä viitekehys kehitetty kuitenkin tutkimaan www-ympäristössä tapahtuvaan profilointia yleisesti eikä keskittymään vain johonkin sen alatekniikkaan. Johtuen kehyksen sisältämien ominaisuuksien yksityiskohtaisesta tasosta se kuitenkin vaatii jatkuvaa päivittämistä pysyäkseen ajantasaisena ja hyödyllisenä.

Tässä luvussa vertailtiin kolmea erilaista profilointimenetelmien vertailun viitekehystä MovieLens-suosittelujärjestelmän avulla. Wun ym. (2002) viitekehyksen todettiin soveltuvan parhaiten personointijärjestelmien tutkimiseen. Middletonin (2003) viitekehys soveltui hyvin suositusjärjestelmien tarkasteluun. Laverin viitekehys taas soveltuu profilointimenetelmien ja –järjestelmien yleiseen tutkimiseen. Viitekehysten ominaisuudet löytyvät liitteestä E. Seuraavassa luvussa suoritetaan yhteenveto tämän työn tuloksista.

6 YHTEENVETO

Tämän konstruktiivisen tutkimuksen tavoitteena oli selvittää, mitä on käyttäjien profilointi www-ympäristössä, mitkä ovat sen ominaisuudet ja kuinka profilointimenetelmiä voidaan vertailla. Tutkimusmenetelmänä käytettiin osittain Nunamakerin, Chenin ja Purdinin (1990) esittämää mallia informaatioteknologian tutkimuksesta. Työssä pyrittiin luomaan jonkinlaista järjestystä käyttäjien profiloinnin sekavaan kenttään niin termien kuin itse profilointimenetelmienkin osalta. Tutkimuksen rajoittajana oli lähinnä vaikeus saada tietoa kaupallisten profilointijärjestelmien toiminnasta. Tärkeimpänä tutkimustuloksena oli profilointimenetelmien ominaisuuksien vertailun viitekehys, joka löytyy liitteestä A.

Luvussa kaksi tehtiin kirjallisuuskatsaus yleisesti käyttäjien profilointiin www-ympäristössä. Luvussa kuvattiin, kuinka profilointi mahdollistaa erottautumisen muista laajalti kilpaillun kentän yrityksistä ja kauppapaikoista. Näitä tapoja olivat henkilökohtainen palvelu, oikeiden tuotteiden tarjoaminen ja yksinkertaisesti niin hyvän asiakassuhteen tarjoaminen, että käyttäjän vaihtokustannukset olisivat kauppapaikkaa vaihdettaessa liian korkeat. Luvussa esiteltiin myös profiloinnin hyödyntämiskeinoja, jotka usein ovat sisällön tai käyttöliittymän personointia. Luvussa paneuduttiin myös muutamiin muihin profilointiin liittyviin kysymyksiin kuten luottamukseen ja käyttäjän tunnistamiseen, joiden molempien todettiin olevan olennainen osa profiloinnin onnistumista.

Seuraavaksi luvussa kolme luotiin kirjallisuuskatsaus olemassa oleviin profilointimenetelmiin. Havaittiin, että on olemassa kahta päälajia profilointia: ihmis- sekä konelähtöistä. Nämä jakaantuvat vielä neljään osaan jotka ovat: yhteistoimintapohjainen, sääntöpohjainen, sisältöpohjainen ja käytönseurantapohjainen profilointi. Tämän lisäksi tutkittiin tiedonsuodatusmenetelmiä, joita tunnistettiin XML-QL sekä kaksi erilaista algoritmia.

Lukujen kaksi ja kolme kirjallisuuskatsausten pohjalta muodostettiin luvussa neljä viitekehys profilointimenetelmien tutkimista varten. Luvussa hahmotettiin yksittäisiä ominaisuuksia edeltävän luvun profiloinnin- ja tiedonsuodatuksen menetelmistä. Näiden perusteella luotiin kuusi pääkohtaa sisältävä viitekehys. Nämä pääkohdat ovat käyttäjän tunnistus, muodostusvaihe, ylläpitovaihe, tiedonsuodatusmenetelmä, kerättyjen tietojen hyödyntäminen sekä käytetyt standardit.

Luvussa viisi verrattiin edellisessä luvussa kehitettyä Laverin viitekehystä kahteen muuhun profilointimenetelmien vertailukehykseen. Laverin viitekehyksen vertailu suoritettiin tutkimalla jokaisen kolmen viitekehyksen avulla MovieLens-suosittelujärjestelmää. Wun ym. (2002) kehittämän viitekehyksen todettiin olevan varsin yleisluontoinen ja kehitetty lähinnä personoinnin tasoa tutkimaan. Middletonin (2002) kehittämän viitekehyksen todettiin olevan suositusjärjestelmien vertailussa hyvä ja tarkka, mutta samalla myös hyvin tekninen. Tässä työssä kehitetty Laverin viitekehys oli joiltain osin vastaava Middletonin kehyksen kanssa. Laverin viitekehys oli yksinkertaisempi termeiltään, laajempi ja näistä kolmesta ainut, joka keskittyy yleisesti profilointimenetelmien vertailuun. Viitekehysten erot ja ominaisuudet löytyvät liitteestä E.

Jatkotutkimusta aiheesta voidaan tehdä esimerkiksi profilointimenetelmien luotettavuuden parissa. Kuinka saada asiakkaat luottamaan kauppapaikkoihin niin paljon, että he suostuvat antamaan henkilökohtaisia tietojaan yrityksen käyttöön? Mahdollisesti ainoa tapa tähän on lisätä asiakkaiden luottamusta www-ympäristöön. Tämä on tietysti käyttäjien profilointia laajempi aihe, mutta parantamalla kauppapaikkansa brändiä voivat yritykset saada asiakkaat luottamaan juuri heihin. Profilointimenetelmien tehokkuutta voidaan myös aina nostaa. Tehokkuutta vaativat suositusjärjestelmien ja personointimenetelmien lisäksi myös profiilin muodostuksen tehokkuus ja tarkkuus. Viimeisenä jatkotutkimuksen aiheena on tässä työssä kehitetty viitekehys, joka vaatii tarkentamista tulevaisuudessa, vastaamaan järjestelmien muuttuvia ominaisuuksia.

Nykyisellä kokoonpanollakaan viitekehys ei välttämättä sisällä kaikkia tarpeellisia ominaisuuksia ja tulevaisuudessa eri menetelmien kehittyessä eteenpäin, se jää ilman jatkotutkimusta jälkeen kehityksestä.

LÄHTEET

Adomavicius, G. ja Tuzhilin, A., 2005. Personalization Technologies: A Process-Oriented Perspective. Communications of the ACM, vol. 48, no. 10, s. 83 – 90.

Allen, C., Kania, D., & Yaeckel, B., (2001). Internet World Guide to One-to-One Web Marketing.

Amazon.com, 2006. Amazon.com. Nähtävillä osoitteessa http://www.amazon.com. [Viitattu 1.11.2006]

Belkin, N. ja Croft , W., 1992. Information Filtering and Information Retrieval: Two Sides of the Same coin? Communications of the ACM, vol. 35 , no. 12, sivut 29 – 38.

Braun, J., 2000. Dempster-Shafer Theory and Bayesian Reasoning in Multisensor Data Fusion. SPIE, vol. 4051, s. 255 – 266.

Carreira, R., Crato, M., Gonçalves, D. ja Jorge, J., 2004. Evaluating Adaptive User Profiles for News Classification. Proceedings of the Ninth International Conference on Intelligent User Interface, s. 206 – 212.

Chellappa, R. ja Sin, R., 2005. Personalization Versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. Information Technology and Management, no. 6, s. 181-202.

Cingil, I., 2002. Supporting Global User Profiles Through Trusted Authorities. ACM SIGMOD Record Archive, vol. 31, no. 1, s. 11 – 17.

Platform for Privacy Preferences (P3P) Project, 2006. Nähtävillä osoitteessa http://www.w3.org/P3P/. [Viitattu 1.11.2006]

Cöner, A., 2003. Personalization and Customization in Financial Portals. Journal of American Academy of Business, vol. 2, no. 2, s. 498 – 504.

Danna, A. ja Gandy, O., 2002. All That Glitters is Not Gold: Digging Beneath the Surface of Data Mining. Journal of Business Ethics no. 40, s. 373 – 386.

Deutsch, A., Fernandez, M., Florescu, D., Levy, A. ja Suciu, D., 1998. XML-QL: A Query Language for XML. Nähtävillä osoitteessa http://www.w3.org/TR/1998/NOTE-xml-ql-19980819/. [Viitattu 1.11.2006]

Egelman, S., Cranor, L. ja Chowdhury, A., 2006. An Analysis of P3P-Enabled Web Sites Among Top20 Search Results. In Proceedings of the Eighth International Conference on Electronic Commerce.

Eirinaki, M. ja Vazirgiannis, M., 2003. Web Mining for Web Personalization. ACM Transactions on Internet Technology, vol. 3, no. 1, s. 1 – 27.

Extensible Markup Language (XML) 1.0 (Fourth Edition), 2006. Nähtävillä osoitteessa http://www.w3.org/TR/2006/REC-xml-20060816/. [Viitattu 1.11.2006]

Friedman, B., Kahn, P. ja Howe, D., 2000. Trust Online. Communications of the ACM, vol. 43, no. 12, s. 34-40.

GroupLens, 2003. GroupLens Reseach. Nähtävillä osoitteessa http://www.grouplens.org/. [Viitattu 1.11.2006]

Herman, L., Swick, R. ja Brickley, D., 2006. Resource Description Framework (RDF). Nähtävillä osoitteessa http://www.w3.org/RDF/. [Viitattu 1.11.2006]

Imhoff, C., Loftis, L., ja Geiger, J., 2001. Building the Customer-Centric Enterprise, Data Warehousing Techniques for Supporting Customer Relationship Management.

Internet World Stats, 2006. World Internet Usage Statistics and Population Stats. Nähtävillä osoitteessa http://www.internetworldstats.com/stats.htm. [Viitattu 1.11.2006]

Middleton, S., 2003. Capturing Knowledge of User Preferences with Recommender Systems. A Thesis for the Degree of Doctor of Philosophy. University of Southampton.

Middleton, S., Shadbolt, N. ja De Roure, D., 2004. Ontological User Profiling in Recommender Systems. ACM Transactions on Information Systems, vol. 22, no. 1, s. 54 – 88.

MovieLens. MovieLens – Movie Recommendations. Nähtävillä osoitteessa http://movielens.umn.edu/. [Viitattu 1.11.2006]

Nunamaker, J., Chen, M. ja Purdin, T., 1990. Systems Development in Information Systems Research. Journal of Management and Information Systems. vol. 7, no. 3. s. 80 – 106.

Pazzani, MJ, 1999. A Framework for Collaborative, Content-Based and Demographic Filtering. Artificial Intelligence Review.

Peppers, D., Rogers, M., ja Dorf, R., 1999. The One to One Fieldbook: The Complete Toolkit for Implementing a 1 to 1 Marketing Program.

Postma, O. ja Brokke, M., 2002. Personalisation in Practice: The Proven Effects of Personalisation. Journal of Database Marketing, no. 9, s. 137 - 142.

Rashid, A., Albert, I., Cosley, D., Lam, S., McNee, S. Konstan, J. ja Riedl, J., 2002. Getting to Know You: Learning New User Preferences in Recommender

Systems. Proceedings of the 2002 International Conference on Intelligent User Interfaces, s. 127 – 134.

Riemer, K. ja Totz, C., 2001. The Many Faces of Personalization – An Integrative Economic Overview of Mass Customization and Personalization. The World Congress on Mass Customization and Personalization.

Risch, D. ja Schubert, P., 2005. Customer Profiles, Personalization and Privacy. Proceedings of the CollECTeR 2005 Conference in Furtwangen.

Sarwar, B., Karypis, G., Konstan, J. ja Reidl, J., 2001. Item-Based Collaborative Filtering Recommendation Algorithms. Proceedings on the Tenth International Conference on World Wide Web, s. 285 – 295.

Schubert, P. ja Leimstoll, U., 2004. Personalization of E-Commerce Applications in SMEs: Conclusions from an Empirical Study in Switzerland. Journal of Electronic Commerce in Organizations, no. 2(3), s. 21 – 39.

Sen, S., Lam, S., Rashid, A., Cosley, D., Frankowsky, D., Osterhouse, J., Harper, M. ja Riedl, J., 2006. Tagging, Communities, Vocabulary, Evolution. Proceedings of Computer Supported Cooperative Work 2006.

Senecal, S. ja Nantel, J., 2004. The Influence of Online Product Recommendations on Consumers' Online Choices. Journal of Retailing 80 , s. 159 – 169.

Serino, C. ja Furner, C., 2005. Making It Personal: How Personalization Affects Trust Over Time. Proceedings of the 38th Hawaii International Conference on System Sciences.

Sobecki, J., 2005. Consensus-Based Hybrid Adaptation of Web Systems User Interface. Journal of Universal Computer Science, vol. 11, no. 2, s. 250 – 270.

Srinivasan, S., Anderson, R. ja Ponnavolu, K., 2002. Customer Loyalty in E-Commerce: An Exploration of Its Antecedents and Consequences. Journal of Retailing, no. 78, s. 41-50.

Sugiyama, K., Hatano, K., Yoshikawa, M., 2004. Adaptive Web Search Based on User Profile Constructed Without Any Effort from Users. Proceedings of Www 2004, s. 675 – 684.

Suomen Laki. Sähköisen Viestinnän Tietosuojalaki 516/2004. Nähtävillä osoitteessa http://www.finlex.fi/fi/laki/alkup/2004/20040516. [Viitattu 1.11.2006]

Swinyard, W. ja Smith, S., 2003. Why People (Don't) Shop Online: A Lifestyle Study of the Internet Consumer. Psychology & Marketing, no. 20(7), s. 567 - 597.

Teltzrow, M. ja Kobsa, A., 2004. Communication of Privacy and Personalization in E-Business. Designing Personalized User Experiences in E-Commerce, s. 315 – 332.

Thuraisingham, B., 2000. A Primer for Understanding and Applying Data Mining. IT Professional, vol. 2, no. 1, s. 28-31.

Vesanen, J. ja Raulas, M., 2006. Building Bridges for Personalization: A Process Model for Marketing. Journal ofIinteractive Marketing, vol. 20, no. 1, s. 5 - 20.

Ward, M. ja Lee, M., 2000. Internet Shopping, Consumer Search and Product Branding. Journal of Product and Brand Management, vol. 9, no. 1, s. 6-20.

Wikipedia, 2006. P3P. Nähtävillä osoitteessa http://en.wikipedia.org/wiki/P3P. [Viitattu 1.11.2006]

Wong, S.K.M. ja Butz, C.J., 2000. A Bayesian Approach to User Profiling in Information Retrieval. Technology Letters.

Wu, D., Tremaine, M., Instone, K. ja Turoff, M., 2002. A Framework for Classifying Personalization Scheme Used on E-Commerce Websites. Proceedings of the 36th Hawaii International Conference on System Sciences. IEEE Computer Society.

Xie, Y. ja Phoha, V., 2001. Web Clustering from Access Log Using Belief Function. Proceedings of the first International Conference on Knowledge capture, s. 202 – 208.

XSL Transformations (XSLT) Version 1.0, 2006. Nähtävillä osoitteessa http://www.w3.org/TR/xslt. [Viitattu 1.11.2006]

Yu, K., Tresp, V. ja Yu, S., 2004. A Nonparametric Hierachical Bayesian Framework for Information Filtering. SIGIR '04, s. 353 – 360.

LIITE A: VIITEKEHYS

Profilointimenetelmien ominaisuuksien arvioinnin viitekehys			
Käyttäjän tunnistus			
	Eväste		
	Sisäänkirjautuminen		
	Profiili kolmannen osapuolen hallussa		
Muodostusvaihe			
	Valmiiden tietokantojen tuonti		
	Ihmislähtöinen profilointi		
		Yleisten tietojen listat	
		Kauppapaikkaan liittyvät listat	
		Objektien arvostelu	
	Konelähtöinen profilointi		
		Sivujen käytön seuranta	
		Hankintojen seuranta	
		Www-lokitiedostojen analysointi	
Ylläpitovaihe			
	Ihmislähtöinen profilointi		
		Yleisten tietojen listat	

		Kauppapaikkaan liittyvät listat	
		Objektien arvostelu	
	Konelähtöinen profilointi		
		Sivujen käytön seuranta	
		Hankintojen seuranta	
		Www-lokitiedostojen analysointi	
Profilointimenetelmän tiedonsuodatus			
	XML-QL		
	Bayesin teoreeman johdannaiset		
	Olettamusfunktion johdannaiset		
Profiilin hyödyntäminen			
	Käyttäjien luokittelu ryhmiin		
	Vertailu käyttäjien kesken (henkilölähtöinen)		
	Vertailu käyttäjien kesken (objektilähtöinen)		
Profilointimenetelmässä käytettävät standardit ja suositukset			
	P3P		
	XML		

LIITE B: MOVIELENS WUN, TREMAINEN, INSTONEN JA TU-ROFFIN VIITEKEHYKSEN MUKAAN

Konelähtöinen personointi	
Ominaisuus	Pisteet
Sisällön personointi hyödyntäen konelähtöistä tietoa (automaattinen elokuvien suosittelu)	**3**
Konelähtöinen profiloinnin hallinta	**-**
Automaattinen linkkien personointi	**-**
Ihmismäinen käyttöliittymä (esim. tervehditään nimeltä)	**1**
Ihmislähtöinen personointi	
Ominaisuus	Pisteet
Sisällön personointi hyödyntäen käyttäjän arviointeja (käyttäjien suorittama elokuvien arvostelu)	**3**
Ihmislähtöinen personoinnin hallinta	**-**
Käyttäjän muokkaama linkkien personointi	**-**
Asiakaskohtainen ulkoasun muokkaaminen	**-**
Yhteensä	**7**

LIITE C: MOVIELENS MIDDLETONIN VIITEKEHYKSEN MUKAAN

Profilointimenetelmä			
	Konelähtöinen		
		- Toiminnan seuraaminen - Tietoa päättelevät heuristiikat	
	Ihmislähtöinen		
		- Käyttäjien palaute - Käyttäjien suorittama ohjelmointi	X
		- Suodatussäännöt - Käyttäjien luomat ryhmät/kategoriat	
Jaettu tieto			
	Käyttäjien palaute		
		- Palaute objekteista - Esimerkit objekteista - Selaushistoria	X
	Järjestelmätieto		
		- Objekti ryhmät/kategoriat - Järjestelmä heuristiikat	X
Profiilin esitys			
	Vektorimalli Selausjäljet Tietämyspohjainen profiili		X
Tietolähde			
	Sisäinen objektitietokanta Tutkitut www-sivut Järjestelmän ulkopuoliset tapahtumat		X
Suositteluteknikka			
	Heuristiikat Samanlaisuuksien yhdistäminen Yhteistoimintapohjainen suodattaminen		X

LIITE D: MOVIELENS LAVERIN VIITEKEHYKSEN MUKAAN

Profilointimenetelmien ominaisuuksien arvioinnin viitekehys			
Käyttäjän tunnistus			
	Eväste		
	Sisäänkirjautuminen		X
	Profiili kolmannen osapuolen hallussa		
Muodostusvaihe			
	Valmiiden tietokantojen tuonti		X
	Ihmislähtöinen profilointi		
		Yleisten tietojen listat	
		Kauppapaikkaan liittyvät listat	
		Objektien arvostelu	X
	Konelähtöinen profilointi		
		Sivujen käytön seuranta	
		Hankintojen seuranta	
		Www-lokitiedostojen analysointi	
Ylläpitovaihe			
	Ihmislähtöinen profilointi		
		Yleisten tietojen listat	

		Kauppapaikkaan liittyvät listat	
		Objektien arvostelu	X
	Konelähtöinen profilointi		
		Sivujen käytön seuranta	
		Hankintojen seuranta	
		Www-lokitiedostojen analysointi	
Profilointimenetelmän tiedonsuodatus			
	XML-QL		
	Bayesin teoreeman johdannaiset		
	Olettamusfunktion johdannaiset		
Profiilin hyödyntäminen			
	Käyttäjien luokittelu ryhmiin		
	Vertailu käyttäjien kesken (henkilölähtöinen)		
	Vertailu käyttäjien kesken (objektilähtöinen)		X
Profilointimenetelmässä käytettävät standardit ja suositukset			
	P3P		
	XML		X

LIITE E: PROFILOINNIN ARVIOINTIVIITEKEHYKSET

	Wun ym.	Middleton	Laveri
Soveltuu	Personointitason tutkimiseen	Suosittelujärjestelmien tutkimiseen	Profilointimenetelmien yleiseen tutkimiseen
Ei sovellu	Yleiseen profilointimenetelmien tutkimiseen	Yleiseen profilointimenetelmien tutkimiseen	Tarkkaan tekniseen analyysiin jonkun tietyn profilointimenetelmän osa-alueen kohdalla
Ominaisuudet	• Tarkasti määritelty ainoastaan erilaisia personoinnin keinoja • Tuloksia voi verrata ainoastaan muihin samalla viitekehyksellä vertailtuihin järjestelmiin tai menetelmiin	• Matemaattinen lähestymistapa • Varsin tekninen • Yksityiskohtainen joiltakin profiloinnin osa-alueilta • Vaatii tarkkaa tutustumista	• Sisältää yleisiä profilointimenetelmien ominaisuuksia • Vaatii tutustumista