

Sara Afflekt-Lazar

# Suosittelujärjestelmät ja niiden tekniikat

Tietotekniikan kandidaatintutkielma

15. toukokuuta 2018

Jyväskylän yliopisto

Informaatioteknologian tiedekunta

**Tekijä:** Sara Afflekt-Lazar

**Yhteystiedot:** sara.k.afflekt-lazar@student.jyu.fi

**Työn nimi:** Suosittelujärjestelmät ja niiden tekniikat

**Title in English:** Recommender Systems and their techniques

**Työ:** Kandidaatintutkielma

**Sivumäärä:** 24+0

**Tiivistelmä:** Suosittelujärjestelmät ovat nykyään yleisiä niin verkkokaupoissa kuin muilla internetsivuistoilla. Tämän kandidaatintutkielman tarkoituksena on tutkia suosittelujärjestelmiä yleisesti, niiden tärkeimpiä tekniikoita sekä sovelluksia joissa suosittelujärjestelmiä käytetään. Suosittelujärjestelmien tekniikoilla on eroja, sekä jokaisella tekniikalla omat etunsa ja haittansa. Yritykset ovatkin näin ollen keksineet omia suosittelujärjestelmien tekniikoita, jotka sopivat heidän tarpeisiinsa parhaiten.

**Avainsanat:** Suosittelujärjestelmät, yhteistoiminnallinen suodatus, hybridimetodit, sisältöperusteinen suodatus, algoritmi.

**Abstract:** Recommender systems are nowadays very common and widely used in net stores and in many websites. Purpose of this bachelor thesis is to examine recommender systems in general, their most important techniques and applications and sites where recommender systems are used. Techniques of recommender systems have their differences, and their own benefits and issues. Many companies have created their own types of recommender systems and techniques, which are most suited for their purposes.

**Keywords:** Recommender systems, collaborative filtering, hybrid methods, content based filtering, algorithm.

## **Kuviot**

Kuvio 1. Esimerkki Netflixin suosittelusta. (Haettu 23.4.2018.) .....	3
Kuvio 2. Käyttäjiin perustuva yhteistoiminnallinen prosessi Badaro ym. (2013) .	7
Kuvio 3. Sisältöperusteisen suosittelujärjestelmän arkkitehtuuri .....	10

## Sisältö

1	JOHDANTO .....	1
2	SUOSITTELUJÄRJESTELMÄT YLEISESTI .....	2
3	TEKNIIKAT JA METODIT .....	5
	3.1 Yhteistoiminnallinen suodatus .....	5
	3.2 Sisältöperusteinen suodatus .....	9
	3.3 Hybridimetodit .....	11
4	SUOSITTELUJÄRJESTELMÄT KÄYTÄNNÖSSÄ.....	13
	4.1 E-kaupankäynnissä .....	13
	4.2 Sosiaalisissa verkostoissa .....	15
	4.3 Online-kirjastoissa.....	16
5	YHTEENVETO .....	17
	KIRJALLISUUTTA .....	18

# 1 Johdanto

Internet ja monet sovellukset sisältävät paljon tietoa, joten tähän informaatiotulvaan on kehitetty suosittelujärjestelmiä, joita käytetään apuna tiedon jakamisessa käyttäjille. Yrityksille suosittelujärjestelmät antavat erinomaisen mahdollisuuden suosittella asiakkaalle tuotteita joista asiakkaat pitävät. Tämä onkin erittäin tärkeää esimerkiksi e-kaupoissa, jonka valikoimissa on tuhansittain tuotteita, jotka eivät kiinnosta asiakasta, mutta suosittelujärjestelmien avulla pystytään suuren tietomäärän keskeltä löytämään asiakasta kiinnostavat tuotteet.

Suosittelujärjestelmien kehittäminen on monialainen projekti johon kuuluu monenlaisia osaajia tekoälyn asiantuntijoista, tiedonlouhijoihin, tilastotieteilijöihin ja markkinoinnin asiantuntijoihin.

Tulen tässä kirjallisuuskatsauksessa perehtymään suosittelujärjestelmiin eri konsepteissa sekä niiden tekniikoihin. Tutkin mihin ja miksi suosittelujärjestelmiä käytetään ja miksi ne ovat tämän päivän internetin valloittamassa maailmassa erittäin hyödyllisiä järjestelmiä niin yrityksille, kuin keskiverto tietojärjestelmien/internetin käyttäjälle. Tulen käsittelemään erilaisia tekniikoita ja niiden hyötyjä ja haittoja, sekä millaisia tekniikoita eri yrityksissä suositaan ja mitä syitä suosittelujärjestelmien käytölle on.

Toisessa luvussa käyn läpi suosittelujärjestelmiä yleisesti ja tutustun perussanastoon liittyen suosittelujärjestelmiin. Kolmannessa luvussa tutkin kolmea yleisintä suosittelujärjestelmän tekniikkaa: yhteistoiminnallista suodatusta, sisältöperustaista suodatusta sekä hybridimetodeja.

Neljännän luvun pääkohtia ovat suosittelujärjestelmät käytännössä. Kirjoitan suosittelujärjestelmistä e-markkinoinnissa, online-kirjastoissa sekä sosiaalisissa verkostoissa. Neljännän luvun keskiössä ovat kolme yritystä, Netflix, Amazon ja Facebook, joiden tekniikoita ja metodeja tutkin.

## 2 Suosittelujärjestelmät yleisesti

Suosittelujärjestelmät ovat ohjelmistotyökaluja ja -tekniikoita tarjoten ehdotuksia kohteista, jotka ovat käyttökelpoisia käyttäjälle.

Suosittelujärjestelmien kehittäminen työllistää monen eri alan osaajia IT-alan asiantuntijoista, data-analyttikoihin, markkinoinnin ammattilaisiin sekä kognitiotieteilijöihin.

Suosittelujärjestelmät pohjautuvat algoritmien käyttöön. Järjestelmien kehittämisessä on kuitenkin huomattu tärkeäksi ottaa huomioon myöskin käyttäjäkokemus sekä ihmisen psykologia.

Burke ja Ramezani (2011) mukaan toisin kuin useat muut ohjelmistojärjestelmät, suosittelujärjestelmät eivät ole määritelty tietyn laskentatavan, tallennustavan tai tietyn tyyppisen datan käytön mukaan. Suosittelujärjestelmillä voikin olla monta erilaista roolia, riippuen siitä mihin palveluntarjoaja haluaa niitä käyttää. Ricci (2011).

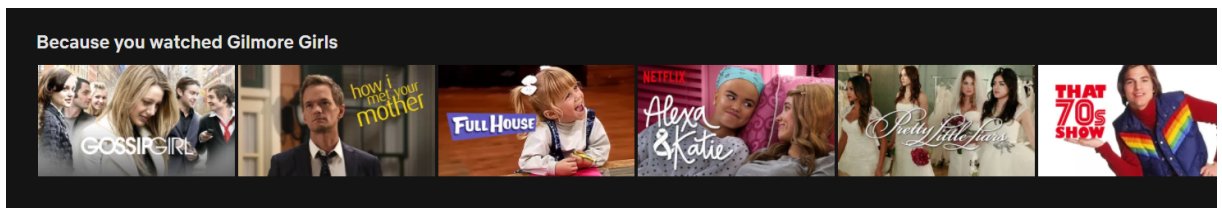
Ricci (2011) mukaan suosittelujärjestelmien tutkimus on varsin uusi tutkimusala verrattaen esimerkiksi muihin klassisiin tietojärjestelmien työkaluihin ja niiden tutkimukseen, kuten esimerkiksi tietokantoihin ja hakumoottoreihin.

Suosittelujärjestelmät luottavat monen tyyppiseen syötteeseen prosessoidessaan dataa. Järjestelmälle kaikista "helpoin" on korkealaatuinen, täsmällinen palaute (engl. explicit feedback) Ricci (2011). Tällaista palautetta on esimerkiksi Netflixin peukalon ylös taikka alas klikkaaminen (entinen tähtiluokittelu) elokuvan katselun jälkeen. Koska täsmällistä palautetta ei ole aina saatavilla, jotkut suosittelijat päättelivät käyttäjän mieltymykset runsaammasta implisiitisestä eli epäsuorasta palautteesta.

Suosittelujärjestelmien kaksi tärkeintä osaa ovat: käyttäjät (engl. users) ja kohteet (engl. items).

Kohteet ovat objekteja joita suositellaan. Kohteet pystytäänkin kuvailemaan niiden monipuolisuuden sekä hyödyn mukaan. Mm. Ricci (2011) mukaan kohteen arvo voi olla positiivinen, jos se on hyödyllinen käyttäjälle, tai negatiivinen jos kohde ei ole sopiva ja käyttäjä teki väärän päätöksen valitessaan kohteen.

Käyttäjillä voi olla monta piirrettä ja tavoitetta. Käyttäjien huomioonottaminen onkin siis suosittelujärjestelmissä tärkeää. Suosittelujärjestelmän käyttöliittymän ulkoasun tulisi olla sellainen josta käyttäjät pitävät. Lü ym. (2012) mukaan käyttäjät arvostavat suosittelun läpinäkyvyyttä, joka tarkoittaa tässä tapauksessa sitä, että käyttäjälle on selvää miksi kyseistä kohdetta on hänelle suositeltu. Esimerkiksi Netflix Netflix (2018) näyttää katsojalle listan suosituksista selittäen "Koska katsoit Gilmoren tyttöjä", ja antaa listan suosituksista jotka liittyvät sarjaan Gilmoren tytöt 1.



Kuvio 1. Esimerkki Netflixin suosittelusta. (Haettu 23.4.2018.)

Koska lista potentiaalisesti mielenkiintoisista kohteista voi olla hyvinkin pitkä, suosittelujärjestelmän täytyy esittää lista mahdollisimman yksinkertaisesti, jotta käyttäjän olisi mahdollisimman helppo selailta erilaisia suosituksia Lü ym. (2012). Edellä mainittu selailun helpotus näkyy hyvin kuviossa 1.

Kolmas osa, joka on tärkeää huomioida suosittelujärjestelmistä puhuttaessa ja niitä kehittäessä, on järjestelmien kohdealue (engl. domain). Burke ja Ramezani (2011) mukaan suosittelun kohdealue on kokoelma kohteita, jotka suosittelujärjestelmä tulee käymään läpi, mutta saattaa myös ottaa mukaan kokoelmia tavoitteita tai tarkoituksia joita suosittelun tulisi tukea. Toinen näkökulma kohdealueesta liittyy suurempaan applikaatioon, jonne suosittelujärjestelmä on sisäänrakennettu Burke ja Ramezani (2011).

Kuten kaikille tietojärjestelmille ja niiden osille, on suosittelujärjestelmillekin tärkeää tulla testatuksi ja arvioiduksi. Jotta suosittelujärjestelmä pystytään arvioimaan kunnolla, on tärkeää ymmärtää järjestelmän tavoitteet ja toiminnot mihin kyseistä järjestelmää aiotaan käyttää. Herlocker J. ym. (2004)

Suosittelujärjestelmien arviointi on vaikeaa useasta eri syystä. Lü ym. (2012) kirjoittavat suosittelualgoritmien vertailun olevan ongelmallista yksinkertaisesti siksi, koska toisistaan erilaiset algoritmit voivat ratkaista täysin erilaisia tehtäviä. Herlocker J. ym. (2004) on listannut alan yleisesti referoituun artikkeliin syitä suosittelujärjestelmän arvioinnin haastavuudelle. Ensinnäkin erilaiset algoritmit voivat olla hyviä taikka huonoja riippuen minkälaisissa joukoissa tietueita nämä toimivat. Toiseksi, tavoite miksi järjestelmää arvioidaan voi myöskin vaihdella. Herlocker J. ym. (2004) mukaan on myöskin huomattavan haastavaa päätettäessä mikä kombinaatio keinoja käytetään arvosteluprosessissa. Arvionnin haastavuuteen vaikuttaa myöskin käyttäjäkokemukset suosittelujärjestelmissä, kirjoittavat Lü ym. (2012). Käyttäjäkokemukseen kuuluvat käyttäjän tyytyväisyys annettuun suositteluun ja käyttäjän luottamus järjestelmään. Näitä kaikkia on vaikea arvioida ns. "offline-ympäristössä.



## 3 Tekniikat ja metodit

Suosittelujärjestelmillä on useita erilaisia tekniikoita joita käytetään. Useissa suosittelujärjestelmissä käytetäänkin nykyään useampaa kuin yhtä tekniikkaa ja metodia Lops, Gemmis ja Semeraro (2011).

Balabanovic ja Shoham (1997) mukaan sisältöperusteinen suosittelujärjestelmä pyrkii suosittelemaan samankaltaisia kohteita, joista annettu käyttäjä on aikaisemmin pitänyt, kun taas suosittelujärjestelmä, suunniteltu yhteistoiminnallisen suosittelu-paradigman mukaan tunnistaa käyttäjät kenen mieltymykset ovat samankaltaisia kuin annetun käyttäjän, ja suosittelevat kohteita joista he ovat pitäneet.

### 3.1 Yhteistoiminnallinen suodatus

Koren ja Bell (2011) mukaan yhteistoiminnallinen suodatus (engl. collaborative filtering) oli ensimmäinen käytetty suosittelujärjestelmien tekniikka, joka tekeekin siitä historiallisesti merkittävän. Netflixin suosio ja sen käyttämät suosittelujärjestelmien tekniikat ovat nostaneet kiinnostusta yhteistoiminnallisuodatukseen viime vuosien aikana.

Yhteistoiminnalliset suodatusmenetelmät tuottavat käyttäjäkohtaisia suosituksia kohteista, jotka perustuvat luokitusten tai käyttötapojen (esim. ostojen) malleihin, ilman ulkoisten tietojen antamista kohteesta tai käyttäjistä Ricci (2011).

Kembellec ym. (2014) kirjoittaakin yhteistoiminnallisten suosittelujärjestelmien pohjautuvan käyttäjien mielipiteiden tilastolliseen prosessointiin.

Tämä tekniikka edellyttää käyttäjien todentamista sisällönhallintaympäristössä. Kembellec ym. (2014) mukaan kun järjestelmä on ehdottanut dokumenttia käyttäjälle pohjautuen kriteereihin, jotka ollaan saavutettu profiilin luonnin aikana ja/tai käyttäjän sisäisen lisähakukoneen käytön aikana, jälkimmäinen tulee ehdottamaan mahdollisuutta luokitella siihen luokitus. Tämä kyseinen luokitus tulee siten pysymään järjestelmässä uudelleen käytettäväksi.

Ricci (2011) kertoo kirjassaan *Recommender Systems Handbook*, että yhteistoiminnallisen lähestymistavan yksinkertaisin ja alkuperäinen toteutus suosittelee aktiiviselle käyttäjälle kohteita, joista muut samankaltaiset käyttäjät pitivät aikaisemmin käyttössään.

Suosituksen laatimiseksi yhteistoiminnallisten suosittelujärjestelmien, kuten useiden muidenkin suosittelujärjestelmien, on yhdistettävä kaksi olennaisesti erilaista kokonaisuutta: kohteet ja käyttäjät.

Yhteistoiminnallisessa suodatuksessa käyttäjät ollaan mallinnettu yksinkertaiseksi listaksi, sisältäen käyttäjän tekemät luokittelut kohteisiin Koren ja Bell (2011).

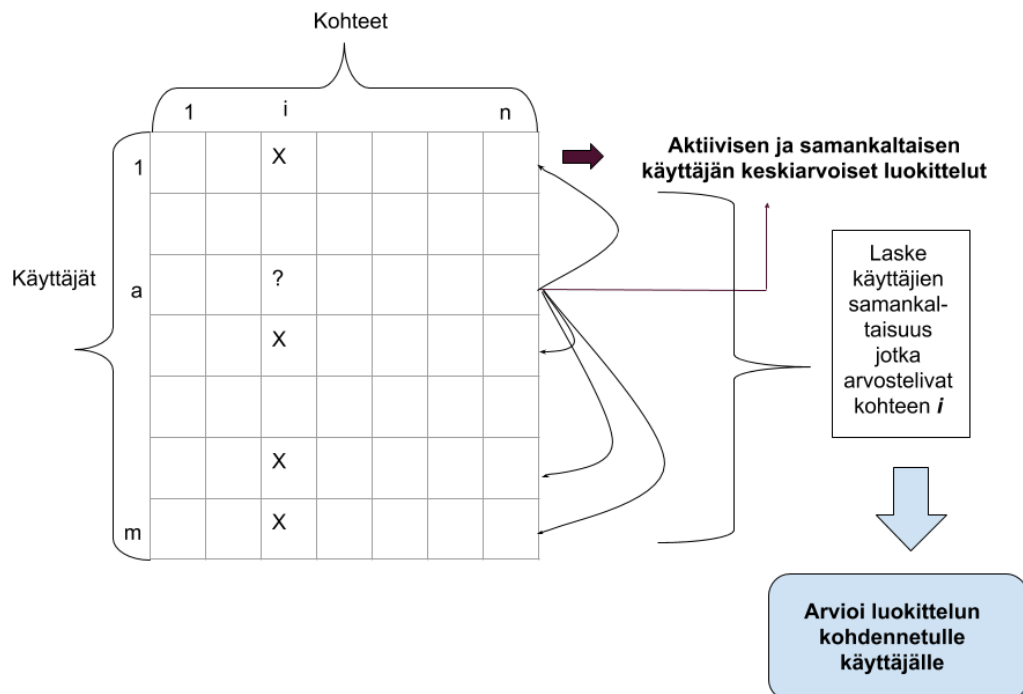
Käyttäjiin perustuvassa yhteistoiminnallisessa tekniikassa käyttäjillä, jotka luokittelevat saman kohteen samoin tavoin, on todennäköisesti sama maku. Drachsler ym. (2008) mukaan tähän oletukseen perustuen tämä tekniikka suosittelee ei vielä nähtyjä kohteita, joita samankaltainen käyttäjä on luokitellut. Käyttäjiin perustuvalla tekniikalla ei ole sisältöanalyysijä, eikä se ole kohdealueellisesti itsenäinen (engl. domain-independent). Tässä tekniikassa laatu parantuu ajan myötä, sekä sillä on kokoava (engl. bottom-up) lähestymistapa Drachsler ym. (2008). Laadun parantuminen ajan myötä sekä kokoava lähestymistapa voidaan laskea tämän tekniikan eduksi/hyödyiksi. Yhteistoiminnallisen suodatuksen eduksi on myös kuvailtu onnekkasta sattumaa. Schani ja Gunawardana (2011) mukaan tämä onnekas sattuma on käytännössä mittari, kuinka yllättävää onnistunut suosittelu käyttäjälle on.

Esimerkin käyttäjiin perustuvasta suosittelun prosessista esittelivät Badaro ym. (2013), josta esimerkki näkyy kuviossa 2.

Kohteeseen perustuvassa tekniikassa fokus on kohteissa, olettaen että kohteet jotka ovat luokitettu samankaltaisesti ovat siten todennäköisesti samankaltaisia Drachsler ym. (2008). Tämä tekniikka suosittelee kohteita, joilla on korkein korrelaatio perustuen käyttäjän dataan luokittelun sijaan.

Kohteiden ja käyttäjien vertailun helpottamiseksi on kaksi ensisijaista lähestymistapaa, jotka muodostavat siten kaksi muuta yhteistoiminnallisen suodatuksen pää-

Kuvio 2. Käyttäjiin perustuva yhteistoiminnallinen prosessi Badaro ym. (2013)



tekniikkaa, joita ovat naapuruusmenetelmät sekä latentit faktorimallit.

Koren ja Bell (2011) kuvailevat, naapuruusmenetelmien keskittyvän kohteiden väliin suhteisiin tai vaihtoehtoisesti suhteisiin käyttäjien välillä, ja näin onkin myös samalla yleisin yhteistoiminnallisuusuosittelujärjestelmien tekniikka.

Latentit faktorimallit (engl. latent factor models), kuten matriisien faktorointi (engl. matrix factorization) käsittävät vaihtoehtoisen lähestymistavan muuttamalla sekä kohteita että käyttäjiä samaan latenttiin faktoritilaan. Ricci (2011) mukaan latentti-tila yrittää selittää luokituksia kuvaamalla sekä tuotteita että käyttäjiä sellaisista tekijöistä, jotka automaattisesti johtuvat käyttäjän palautteesta.

Matriisien faktorointi tarkoittaa prosessia, jossa matriisi on faktoroitu tulokseen monia matriiseja Symeonidis (2017). Symeonidis (2017) muistuttaakin että yksi matriisien faktoroinnin vahva puoli on se, että se sallii lisätietojen sisällyttämisen. Li-

sätietojen sisällyttäminen on tärkeää, sillä kun eksplisiittistä tietoa ei ole saatavilla, suosittelujärjestelmä voi päätellä käyttäjän mieltymykset käyttäen implisiittistä tietoa. Tämä taas vaikuttaa epäsuorasti mielipiteisiin tarkkaillessaan käyttäjän käyttäytymistä ottaen lukuun mm. ostokset, selaushistorian hakumallit, sekä hiiren liikeradat Symeonidis (2017) .

Yhteistoiminnalliseen suodatukseen liittyy myös tietty määrä ongelmia, joita on mm. kylmä alkua, skaalautuvuus sekä aikaisen arvostelijan ongelma. Ongelmia yhteistoiminnallisesta suodatuksesta on listannut useat eri kirjoittajat mm. Valois et al. (2011) ja Drachsler ym. (2008)

Kylmän alun ongelma (engl. cold start problem) esiintyy järjestelmän alussa. Valois et al. (2011) mukaan tämä ongelma vaatii käyttäjää omistautumaan oman profiilin luomiseen ennen kuin järjestelmä kykenee aloittamaan yhtään merkityksellistä suosittelua. Tämän vuoksi kohteet joita ei ole aikaisemmin arvioitu, jätettäisiin kokonaan huomiotta. Kylmän alun ongelma esiintyy niin kohteeseen perustuvassa, että käyttäjiin perustuvassa tekniikassa.

Jos käyttäjällä on suosittu/yleinen maku, voi se koitua ongelmaksi yhteistoiminnallisessa suodatuksessa, sillä mitä enemmän/suosittu kohde on, sitä enemmän sitä tullaan suosittelemaan, vaikeuttaen potentiaalisten vähemmän tunnetuiden tuotteiden suosittelua oikealle käyttäjälle.

Skaalautuvuus on ongelmana käyttäjiin perustuvassa tekniikassa. Valois et al. (2011) mukaan kun käyttäjien määrä, kohteet ja arvioinnit ovat liian suuria, järjestelmä joka toteuttaa reaaliaikaiset laskutoimitukset käyttäjien välisistä suhteista saattaa antaa hyvin pitkän vastausajan, jonka seurauksena saattaa tarvita tietokoneen resursseja jotka eivät ole silloin saatavilla.

Valois et al. (2011) mukaan niin sanottu aikaisen arvostelijan ongelma (engl. early rater issue), tapahtuu kun uusi kohde ilman käyttäjän arvostelua tai suositusta lisätään ja sitä ei voida vielä suositella.

Muita ongelmia ovat esitelleet mm. Drachsler ym. (2008), joita ovat kummallekin

yhteistoiminnalliselle suodatuksen tekniikalle uusi-käyttäjä ongelma, uusi-kohde ongelma sekä niukkuus.

### 3.2 Sisältöperusteinen suodatus

Sisältöperusteinen suodatus (eng. content based filtering) on toinen klassinen suosittelujärjestelmien käyttämä tekniikka, joka perustuu järjestelmän suositteleman sisällön kuvaukseen ja analyysiin. Kembellec ym. (2014) mukaan tämä prosessi perustuu suuresti tekstin analysoinnin tekniikoihin, mutta voidaan kuitenkin laajentaa metatietoihin sisältäviin sisältöihin.

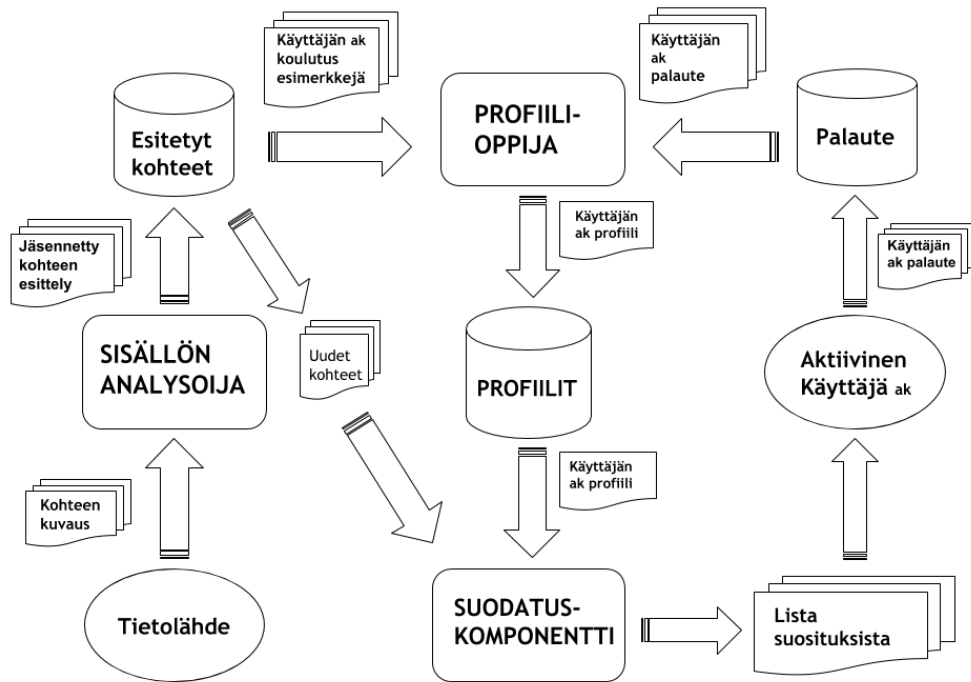
Ma, M. Y. ja Uchyigit, G. (2008) kirjoittavat sisältöperusteisen suosittelun koostuvan suosituksista käyttäjälle, joista käyttäjä on pitänyt ennen.

Burke ja Ramezani (2011) kirjoittavat sisältöperusteisensuosittelun olevan yksilöllisesti keskittynyt, käyttäen kohteen ominaisuuksia ja käyttäjän mielipiteitä kouluttaakseen lajittelukoneen, joka voi ennustaa mieltymyksiä uusista kohteista. Jotkut lähestymistavat perustuvat kuitenkin yksinkertaiselle vertailulle avainsanojen välillä Ma, M. Y. ja Uchyigit, G. (2008). Ricci (2011) mukaan järjestelmä oppii suosittelemaan kohteita, jotka ovat samankaltaisia mistä käyttäjä on pitänyt aikaisemmin. Kohteiden samankaltaisuus lasketaan ominaisuuksien pohjalta, jotka ovat liitettävissä verrattavien kohteiden kanssa.

Tietyt sisältöperusteisten suodatusten toteutukset luottavat automaattisten lajittelukoneiden ennustavaan kapasiteettiin, joita ovat esimerkiksi neuroverkot, puudiaagrammit, päättelysäännöt sekä bayesilaiset verkot Ma, M. Y. ja Uchyigit, G. (2008). Edellä mainitut ovat laskennallisia malleja jotka luokittelevat annetun syötteen spesifinä kategoriana ennaltamäärätyinä koulutusarjana Ma, M. Y. ja Uchyigit, G. (2008).

Järjestelmät pohjautuen sisältöperusteiseen suositusmenetelmään analysoivat siis joukon dokumentteja ja/tai kuvauksia, joita käyttäjä on aiemmin arvostellut, ja sitten rakentavat mallin tai profiilin käyttäjän kiinnostuksista perustuen kyseisen käyt-

Kuvio 3. Sisältöperusteisen suosittelujärjestelmän arkkitehtuuri



täjän objektien ominaisuuksiin Mladenic (1999). Ricci (2011) kuvailee profiilia jäsennehtynä esityksenä käyttäjien kiinnostusten kohteista ja kertoo kuinka järjestelmä omaksuu siten suosittelemaan uusia kiinnostavia kohteita. Ma, M. Y. ja Uchyigit, G. (2008) kertovat sisältöperusteisen metodin tarvitsevan mittajärjestelmän määrittääkseen samankaltaisuudet käyttäjien profiilien ja kohdediagrammin välillä.

Sisältöperusteisen suosittelujärjestelmän arkkitehtuuria on esitelty kuviossa 3, joka perustuu kuvioon, jonka esittivät Lops, Gemmis ja Semeraro (2011) s.76.

Sisältöperusteisella suodatuksella on omia tekniikoita, joihin kuuluvat tapauskohtainen ajattelu (engl. case-based reasoning) sekä ominaisuuksiin pohjautuva tekniikat (engl. attribute based techniques).

Tapauskohtainen ajattelu olettaa että jos käyttäjä pitää tietystä kohteesta, hän aikoo todennäköisesti pitää myöskin samankaltaisista kohteista.

Tapauskohtaisen ajattelun hyötyihin lukeutuvat kohdealueen itsenäisyys (engl. domain-

independence), sekä että laatu parantuu ajan myötä. Drachsler ym. (2008) Tapauskohtainen ajattelu ei sisällä sisältöperusteista analyysia joten tämänkin Drachsler ym. (2008) laskevat tekniikan hyödyiksi. Drachsler ym. (2008) mukaan, tapauskohtaisen tekniikan haittoja ovat: "uusi-käyttäjä ongelma", ylierikoistuminen, niukkuus ja kylmä alkua (engl.cold start problem).

Ominaisuuksiin pohjautuva tekniikka suosittelee kohteita, pohjautuen attribuuttien sovittamiseen käyttäjän profiiliin. Attribuutit voivat olla painotettu niiden eri merkityksestä käyttäjälle Drachsler ym. (2008).

Yksi sisältöperusteisen suodatuksen ominaisuuksiin pohjautuvan tekniikan hyötyihin kuuluu ongelman "kylmästä alusta"puuttuminen Drachsler ym. (2008), josta esimerkiksi yhteistoiminnallinen suodatus kärsii. Tämä on yksi syy miksi sisältöperusteinen suodatus voi toimia kilpailukykyisempänä yhteistoiminnalliseen suotukseen verrattuna.

### **3.3 Hybridimetodit**

Hybridimetodit perustuvat yhdistelmään yllämainituista tekniikoista. Hybridijärjestelmän ottaa jokaisesta tekniikasta jotain ja yhdistää ne.

Yleisin hybridisuosittelevä järjestelmä tulee kun yhdistetään yhteistoiminnallinen suodatus ja sisältöperusteinen suodatus.

Ricci (2011) esittää yksinkertaisen havainnollistavan esimerkin: Hybridijärjestelmä yhdistäen tekniikoita A ja B, ottaa se esimerkiksi A:n hyödyt ja käyttää näitä hyötyjä korjatakseen B:n epäkohdat.

Hybridimetodit ovatkin kaikista tehokkaimpia suosittelujärjestelmien metodeja, perustuen sen kykyyn yhdistää parhaat tekniikat muista suosittelujärjestelmien tekniikoista.

Robert Burke (2002) on esitellyt 7 hybridisaatiotekniikkaa, joita ovat painotettu, vaihtelu, sekoitettu, ominaisuuksien yhdistäminen, vesiputous, ominaisuuksien li-

sääminen sekä meta-taso.

Painotetussa metodissa kohteen suosittelun laatu perustuu käytettävien metodien summaan. Vaihtelussa järjestelmä valitsee käytettäväksi joko dataan perustuvan metodin, taikka sosiaalisen suodatuksen riippuen käyttäjän hakukontekstista. Kembellec ym. (2014) mukaan ominaisuuksien yhdistäminen tarjoaa mahdollisuuden rikastuttaa dataa mikä on integroitu "edeltävään" järjestelmään käyttäjien luokitusten kanssa, joka siten rikastuttaa tietokantaa "jälkikäteen". Suosittelun laskelma toteutetaan siten koko datan laajuisesti.

Sekoitettu teknologia helpottaa suosittelujen ehdotelmaa perinteisistä metodeista tarkoituksenaan rajoittaa klassisten menetelmien haitat. Vesiputousprosessi hybridisuosittelujärjestelmien kontekstissa koostuu käyttäjäprofiilien tupla-analyysistä Kembellec ym. (2014). Ensimmäinen osa on käytetty korostamaan potentiaalisia ehdokkaita. Toinen osa on käytetty taasen tarkentamaan käyttäjävalikoimaa.

Ominaisuuksien lisääminen toimii samankaltaisesti vesiputouksen kanssa. Kuitenkin jos ehdokkaiden määrä on liian suuri ensimmäisessä osassa, toinen osa tulee toteuttamaan sitten toissijaisen erottelun/arvostelun integroiden suositteluiden tuotteiden dataan Kembellec ym. (2014). Meta-tasossa suodatetaan käyttäjät kahdesti, jotta samankaltaisuudet pystytään määrittämään Kembellec ym. (2014). Tämä metodi on samankaltainen ominaisuuksien lisäämisen sekä vesiputouksen kanssa.



## 4 Suosittelujärjestelmät käytännössä

Suosittelujärjestelmiä käytetään monenlaisissa sovelluksissa ja verkkosivuilla. Ricci (2011) kirjoittaaakin suosittelujärjestelmien olevan tärkeässä roolissa monissa korkeasti arvosteluilla verkkosivuilla, kuten Amazon.com (2018) ja Netflix (2018). Silloin kun saatavilla on yllin kyllin monenlaisia tavaroita ja asiakkaita, suosittelujärjestelmät pystyvät varmasti asiakasta selailussa auttamaan. Lü ym. (2012) kirjoittavat suosittelujärjestelmällä olevan huomattava vaikutus yrityksen liikevaihtoon. Esimerkin on esittänyt tästä Jon Sanders puheessaan "Research Challenges in Recommenders" kolmannessa ACM konferenssissa suosittelujärjestelmistä (2009), jonka Lü ym. (2012) ovat artikkelissaan esittäneet: Jopa 60 prosenttia DVD:istä joita on myyty Netflixin kautta, on valittu juurikin suosittelujärjestelmän esitettyjen suositusten kautta.

### 4.1 E-kaupankäynnissä

Suosittelujärjestelmiä käytetään laajasti e-kaupankäynnissä kohdennettuna markkinoinnin työkaluna.

Lü ym. (2012) mukaan suosittelujärjestelmät eivät vain auta siinä mitä tuotteita yhdelle asiakkaalle tarjotaan, vaan ne myös parantavat yrityksen/verkkokaupan myyntiä ehdottelemalla extratuotteita asiakkaalle ja parantavat näin myös asiakkaan uskollisuutta yritystä kohtaan, sillä kuluttajilla on tapana palata sivustoille jotka palvelevat heidän tarpeitaan parhaiten.

E-kaupankäynnin algoritmit toimivat yleensä haastavissa ympäristöissä. Uusien asiakkaiden tapauksessa heistä on hyvin rajattu määrä tietoa, joka perustuu pieniin määriin ostoksia taikka arvosteluja. Taasen vanhemmalla asiakkaalla saattaa olla hyvin laajasti erilaista tietoa perustuen moniin ostoksiin sekä luokitteluihin.

Isolla jälleenmyyjällä saattaa olla valtava määrä dataa, miljoonittain asiakkaita sekä lueteltuja kohteita, joka tuo useanlaisia haasteita erityisesti skaalautuvuudelle,

jonka toteavat myös Linden, Smith ja York (2003).

Lü ym. (2012) mukaan, koska mikään suosittelumetodi ei toimii kaikille asiakkaille samaan aikaan parhaiten, ovat monet suuret sivustot tukeutuneet useisiin tiettyihin suosittelutekniikoihin .

Amazon, joka on yksi suurimmista maailmanlaajuisista e-markkinoinnin yrityksistä, käyttää suosittelujärjestelmiä monissa kampanjoissaan sekä useimmissa verkkosivuistaan. Smith ja Linden (2017) mukaan jokainen ihminen joka tulee Amazon.comiin näkee kaupan eri tavoin, koska sivut ovat jokaiselle käyttäjälle henkilökohtaisesti yksilöity perustuen heidän kiinnostustensa kohteisiin. Kauppa muuttuu radikaalistikin jokaiselle käyttäjälle yksilöidyksi, perustuen käyttäjän kiinnostuksen kohteisiin, näyttäen esimerkiksi maalarille maalaustarvikkeita, ihonhoidosta kiinnostuneelle ihonhoitotuotteita ja viherpeukalolle kasveja. Amazon julkaisi kohteeseen perustuvan yhteistoimintasuodatuksen, joka auttoi suositteluissa, silloin ennennäkemättömällä skaalalla Linden, Jacobi ja Benson (1998). Linden, Smith ja York (2003) kirjoittavat artikkelissaan enemmän Amazonin suosittelujärjestelmästä, ja millaisia algoritmeja se käyttää.

Amazon käyttää siis omaa "kohde kohteeseen"(engl. item-to-item) yhteistoiminnallista suodatusta, joka skaalautuu hyvin massiivisille tietueille sopivaksi ja tuottaa korkealaatuisia suosituksia reaaliajassa Linden, Smith ja York (2003). Linden, Smith ja York (2003) mukaan, toisin kuin perinteinen yhteistoiminnallinen suodatus, Amazonin algoritmien online-laskenta skaalautuu itsenäisesti asiakkaiden ja kohteiden eli tuotteiden lukumäärään tuoteluottelossa. Sen sijaan että yhdistettäisiin käyttäjä samankaltaisten käyttäjien kanssa, Amazonin kohde-kohteeseen - suodatus yhdistääkin siten kaikki käyttäjän ostokset sekä arvostelleet tuotteet samankaltaisiin tuotteisiin/kohteisiin, ja sen jälkeen yhdistää nämä sanakaltaiset kohteet listaan suosituksista Linden, Jacobi ja Benson (1998).

## 4.2 Sosiaalisissa verkostoissa

Sosiaalisissa verkostoissa suosittelujärjestelmiä käytetään nykypäivänä laajasti. Pazos Arias (2012) kuvailee sosiaalisen verkoston perustuvan kiinnostuksen kohteiden, aktiviteettien ja sisältöjen jakamiseen. Kuten e-kaupankäynnissäkin, suosittelujärjestelmiä käytetään sosiaalisissa verkostoissa joko innostaakseen käyttäjiä laajentamaan henkilökohtaista sosiaalista graafiaan, tai sitten markkinointiin Symeonidis (2017).

Sosiaaliset verkostot kuten LinkedIn LinkedIn (2018) ja Facebook Facebook (2018) käyttävät suosittelujärjestelmiä ehdotellakseen käyttäjälle uusia kontakteja Valois et al. (2011). Kembellec ym. (2014) mukaan sosiaalisen suosittelun johtaminen on hankalaa brändeille, johon liittyviin ongelmiin suosittelujärjestelmät auttavatkin.

Sosiaalinen suosittelu perustuu siihen että käyttäjälle suositellaan kohteita joista käyttäjän kontaktitkin ovat tykänneet. Esimerkiksi Facebookissa käyttäjälle suositellaan tapahtumia, jonne hänen "kaverinsa" ovat menossa tai joista he ovat kiinnostuneita. Nämä suosittelujärjestelmät generoivat ennusteita eli suosituksia, jotka perustuvat tietoon käyttäjien profiileista sekä suhteista käyttäjien välillä Victor (2011).

Kaikista tunnetuin suosittelija käyttäen sosiaalisia verkostoja suositellakseen muita sosiaalisia kokonaisuuksia kuten kavereita on varmastikin Facebook. Tärkein Facebookin lähestymistapa suositteluun on FoF- algoritmi (kaverin kaveri algoritmi, engl. friend-of-friend algorithm) Pazos Arias (2012).

Pazos Arias (2012) kertoo FoF-algoritmin laskevan kaikki käyttäjän kavereiden kaverit ja suosittelee niitä jotka eivät ole jo käyttäjän kaverilistalla. "Se, kuinka paljon yhteisiä kavereita käyttäjillä on, sitä korkeammin luokiteltu tämä kaverin kaveri on" Pazos Arias (2012). Oletus tämän lähestymistavan takana on, että käyttäjä voisi todennäköisesti pitää ihmisistä, jotka ovat hänen kavereidensa kavereita.

### 4.3 Online-kirjastoissa

Netflixin suosio teki online-videokirjastoista ilmiön, sekä suosittelujärjestelmistä suosituimpia. Erityisesti Netflix Prize -kilpailu, joka alkoi lokakuussa 2006, vahvisti suosittelujärjestelmien edistystä, jonka totesivat mm. Koren ja Bell (2011). Koren ja Bell (2011) mukaan Netflix Prize (2018) kilpailun luonne on rohkaissut erityisen nopeaa kehitystä, jossa innovaattorit rakensivat jokaisen kehittyneen teknologian päälle uutta, parantaakseen ennustusten tarkkuutta.

Gomez-Uribe ym. (2015) mukaan Netflixin suosittelujärjestelmä ei ole vain yksi algoritmi, vaan ennemminkin kokoelma erilaisia algoritmeja tarjoillen erilaisia käyttötapauksia, jotka tulevat yhteen. Jokainen algoritmi Netflixin suosittelujärjestelmässä tukeutuu tilastotieteelliseen ja koneoppimisen tekniikoihin, joihin kuuluvat esimerkiksi valvottu luokittelu (engl. supervised classification) ja regressio Gomez-Uribe ym. (2015).

## 5 Yhteenveto

Suosittelujärjestelmät ovat ohjelmistotyökaluja joiden tarkoitus on suositella käyttäjille heitä kiinnostavia suosituksia.

Erilaisissa suosittelujärjestelmien tekniikoissa on omat haittansa että hyötynsä. Yhteistoiminnallinen ja sisältöperusteinen suodatus on klassisimpia sekä sovelletuimpia suosittelujärjestelmien tekniikoita. Hybridimetodit ovat yleensä sekoitus sisältöperusteistesta että yhteistoiminnallisesta suodatuksesta.

Suosittelujärjestelmien käyttötarkoitus vaikuttaa siihen, kuinka järjestelmää kehitetään, millainen tekniikka otetaan käyttöön, sekä kuinka järjestelmää arvioidaan.

Yhteistoiminnallisen suodatuksen ollessa ensimmäinen suosittelujärjestelmien tekniikka, ei se välttämättä yksinkertaisuudessaan ole tehokkain kaikille järjestelmille.

Yritykset ovat kehitelleet omia algoritmeja ja suosittelujärjestelmiä kehittääkseen sovelluksensa hyötyjä. Amazon käyttää suosittelualgoritmeja laajasti yksilöidäkseen verkkosivunsa jokaisen käyttäjän henkilökohtaisille kiinnostuksille sopivaksi.

Suosittelualgoritmit tuottavat näin ollen tehokkaan muodon kohdennetusta markkinoinnista luoden yksilöllisiä kokemuksia käyttäjille.

Suosittelujärjestelmiä käytetään nykyään laajasti miltei kaikkialla paljon tietoa sisältävissä sovelluksissa ja sivustoilla. Aina järjestelmän ja tekniikan valitsemisessa ei vaikuta vain markkinallinen hyöty, vaan käyttäjäystävällisyyttä pidetään monesti yhtä tärkeänä.

## Kirjallisuutta

- Amazon, Haettu 20.4.2018. *Amazon.com* Haettu osoitteesta <https://www.amazon.com/>
- Aggarwal, C. (2016) *Recommender Systems: The Textbook*. Cham: Springer International Publishing.
- Badaro, Gilbert., Hajj, Hazem., El-Hajj, Wassim., Nachman, Lama. (2013). *A Hybrid Approach with Collaborative Filtering for Recommender Systems*. 2013 9th International Wireless Communications and Mobile Computing Conference, IWCMC 2013. 349-354. 10.1109/IWCMC.2013.6583584.
- Balabanovic, M., Shoham, Y. 1997 *Content-based, Collaborative Recommendation*. Communications of the ACM 40(3), 66–72 (1997)
- Burke R., (2002) *Hybrid recommender systems: survey and experiments*. User Modeling and User-Adapted Interaction , vol. 12, no. 4, pp. 331– 370, 2002.
- Burke R., Ramezani M. (2011) *Matching Recommendation Technologies and Domains* In: Ricci F., Rokach L., Shapira B., Kantor P. (eds) *Recommender Systems Handbook*. Springer, Boston, MA
- Drachler H., Hummel H. G. K., Koper R. (2008). *Personal recommender systems for learners in lifelong learning: requirements, techniques and model*. International Journal of Learning Technology 3(4), 404–423 (2008a).
- Facebook (2018) Haettu 20.4.2018. *Facebook* Haettu osoitteesta <https://facebook.com>
- Gomez-Uribe, Carlos A. ja Hunt, Neil. (2015). *The Netflix Recommender System: Algorithms, Business Value, and Innovation*. ACM Trans. Manage. Inf. Syst. 6, 4, Article 13 (December 2015), 19 pages. DOI:<https://doi.org/10.1145/2843948>
- Herlocker, Jonathan L. Konstan, Joseph A. Terveen, Loren G. and Riedl, John T. (2004). *Evaluating collaborative filtering recommender systems*. ACM Trans. Inf. Syst. 22, 1 (January 2004), 5-53. DOI: <http://dx.doi.org/10.1145/963770.963772>
- Kembellec, Gérald, Ghislaine Chartron, and Imad Saleh. (2014). *Recommender Systems*. London, [England] ; Hoboken, New Jersey: ISTE Limited.

- Koren Y., Bell R. (2011) *Advances in Collaborative Filtering*. In: Ricci F., Rokach L., Shapira B., Kantor P. (eds) *Recommender Systems Handbook*. Springer, Boston, MA.
- Linden, G., Jacobi, J. ja Benson, E. (1998). *Collaborative recommendations using item-to-item similarity mappings*. patent no. US 6.266.649.)
- Linden, G. Smith, B. and York, J. (2003). *Amazon.com recommendations: item-to-item collaborative filtering*. *IEEE Internet Computing*, vol. 7, no. 1, pp. 76-80, Jan/Feb 2003.
- Linkdedin (2018) Haettu 20.4.2018. *Linkdedin* Haettu osoitteesta <https://fi.linkedin.com/>
- Lops P., de Gemmis M., Semeraro G. (2011). *Content-based Recommender Systems: State of the Art and Trends*. In: Ricci F., Rokach L., Shapira B., Kantor P. (eds) *Recommender Systems Handbook*. Springer, Boston, MA.
- Lü, L., Medo, M., Yeung, C., Zhang, Y., Zhang, Z., Zhou, T. (2012). *Recommender systems*. *Physics Reports*. Volume 519, Issue 1, Pages 1-49, DOI: <https://doi.org/10.1016/j.physrep.2012.02.006>.
- Ma, M. Y. ja Uchyigit, G. (2008). *Personalization techniques and recommender systems*. Hackensack, NJ: World Scientific.
- Mladenic, D. (1999). *Text-learning and Related Intelligent Agents: A Survey*. *IEEE Intelligent Systems* 14(4), 44–54.
- Netflix.com, Haettu 20.4.2018. Haettu osoitteesta <https://Netflix.com>
- Netflixprize, Haettu 20.4.2018. Haettu osoitteesta <https://Netflixprize.com>
- Pazos Arias, J. J. k. (2012). *Recommender Systems for the Social Web*. Heidelberg: Springer Berlin Heidelberg
- Ricci, Francesco. toimittaja, Lior. toimittaja Rokach, Bracha. toimittaja Shapira, and Paul B. toimittaja Kantor. (2011). *Recommender Systems Handbook*. 1. Boston, MA: Springer US.
- Schafer, B. (2006). *The Application of Data Mining to Recommender Systems*. *Encyclopedia of Data Warehousing and Mining*, s. 44–47.
- Schani, G., Gunawardana A. (2011). *Evaluating Recommendation Systems*. In: Ricci F., Rokach L., Shapira B., Kantor P. (eds) *Recommender Systems Handbook*. Spring-

ger, Boston, MA

Smith, B., Linden, G. 2017 "*Two Decades of Recommender Systems at Amazon.com*,". in IEEE Internet Computing, vol. 21, no. 3, pp. 12-18, 2017. doi:10.1109/MIC.2017.72

Symeonidis, Panagiotis. (2017). *Matrix and Tensor Factorization Techniques for Recommender Systems*.

Valois B. Jr, Cleomar, ja Oliveira, Marcius Armada de. (2011). *Recommender systems in social networks*. JISTEM - Journal of Information Systems and Technology Management, 8(3), 681-716. DOI: <https://dx.doi.org/10.4301/S1807-17752011000300009>

Victor, P. K. (2011). *Trust Networks for Recommender Systems*. Paris: Atlantis Press.