

# **Relevanssipalautteen toimivuus lyhyillä, eritasoisesti onnistuneilla aloituskyselyillä**

Jari Friman

Tampereen yliopisto  
Informaatiotutkimuksen laitos  
Pro gradu -tutkielma  
Syyskuu 2008

## TIIVISTELMÄ

TAMPEREEN YLIOPISTO

Informaatiotutkimuksen laitos

FRIMAN, JARI: Relevanssipalautteen toimivuus lyhyillä, eritasoisesti onnistuneilla aloituskyselyillä

Pro gradu -tutkielma 63, s., 4 liites.

Informaatiotutkimus

Syyskuu 2008

---

Tutkimuksen tarkoituksena on selvittää kyselyn laajentamisen vaikutusta lyhyisiin, eritasoisesti toimiviin aloituskyselyihin. Kyselyn laajennus on toteutettu relevanssipalautteella ja pseudorelevanssipalautteella. Relevanssipalautteessa käyttäjä valitsee tuloslistasta relevantiksi arvioimansa dokumentin tai useita dokumentteja ja hakujärjestelmä muodostaa dokumenteissa esiintyviä sanoja hyödyntäen muokatun kyselyn. Pseudorelevanssipalautteessa kone poimii automaattisesti tuloslistan kärjestä halutun määrän dokumentteja, joista se edelleen valitsee hakutermejä ja tekee näillä uuden kyselyn.

Päätutkimuskysymykseni on: Kuinka käyttäjän tekemän kyselyn pitäisi onnistua, jotta relevanssipalautteella ja pseudorelevanssipalautteella voitaisiin parantaa hakutehokkuutta? Lisäksi tarkoitukseni on selvittää tarkemmin palautedokumenttien määrän vaikutusta hakutehokkuuteen, sekä huomioida myös relevanssipalautteeseen valittavien dokumenttien relevanssitason vaikutus.

Tutkimus on toteutettu osittaistämävällä Lemur Indri -tiedonhaku-järjestelmällä. Testikokoelmana on TRECUTA -dokumenttikokoelma ja käytössä on 35 hakuaihetta. Hakutehokkuuden mittarina käytetään keskitarkkuuksien keskiarvoa (MAP = mean average precision) ja tulosten tilastollinen merkitsevyys on testattu Friedmanin testillä. Lisäksi tulosten käytännön merkittävyyttä on arvioitu Sparck-Jonesin prosenttiyksiköiden eroon perustuvalla mittarilla.

Lähtötilanteena tutkimuksessa on kolmella eri tasolla onnistuneet, enintään kolmen sanan mittaiset kyselyt, joita kutsutaan tässä heikosti, hieman paremmin ja hyvin onnistuneiksi aloituskyselyiksi. Eritasoiset kyselyt olen muokannut tutkimusta varten hakuaiheiden otsikoissa ja relevanteissa dokumenteissa esiintyvien termien avulla. Tutkimuksessa ei ole käytetty koehenkilöitä.

Tutkimuksen mukaan relevanssipalautteella voidaan parantaa hakutehokkuutta kyselyn onnistuneisuudesta riippumatta. Relevanssipalaute hyvin onnistuneesta aloituskyselystä tuotti enemmän relevantteja dokumentteja kuin heikkotasoisesta kyselystä. Suhteellisesti laskettuna heikkotasoinen kysely hyötyy kuitenkin eniten relevanssipalautteesta. Relevanssipalautteeseen kannattaa tämän tutkimuksen perusteella valita mahdollisimman monta relevanttia dokumenttia relevanssitasosta riippumatta, vaikka hyviin tuloksiin päästään jo yhdelläkin palautedokumentilla.

Tämän tutkimuksen perusteella pseudorelevanssipalaute ei yleensä paranna tai heikennä hakutehokkuutta niin, että ero olisi tilastollisesti merkitsevä. Toisin sanoen vaikutus näyttää olevan hyvin sattumanvarainen. Myöskään kyselyn onnistuneisuudella tai palautedokumenttien määrällä ei näyttänyt olevan vaikutusta pseudorelevanssipalautteen toimivuuteen.

Avainsanat: kyselyn laajentaminen, relevanssipalaute, pseudorelevanssipalaute, kyselyn onnistuneisuus

|   |    |
|---|----|
| <b>1. Johdanto</b> .....  | 1  |
| <b>2. Tiedonhaun tutkimus</b> .....                               | 4  |
| 2.1 Tiedonhaku ja tiedonhakujärjestelmät .....                    | 4  |
| 2.2 Täsmäytysmenetelmät.....                                      | 5  |
| 2.3 Relevanssin käsitteestä.....                                  | 5  |
| 2.4 Relevanssiarvioinneista tiedonhaun tutkimuksessa .....        | 6  |
| <b>3. Kyselyn laajentaminen</b> .....                             | 7  |
| 3.1 Kyselyn laajentamisen malli .....                             | 7  |
| 3.2 Relevanssipalaute.....  | 9  |
| 3.3 Pseudorelevanssipalaute.....                                  | 10 |
| <b>4. Relevanssipalautetekniikoita</b> .....                      | 12 |
| 4.1 Relevanssipalaute vektorimallissa .....                       | 12 |
| 4.2 Relevanssipalaute todennäköisyysmallissa.....                 | 14 |
| 4.3 Avainten valinta ja painottaminen .....                       | 18 |
| 4.3.1 Kwokin (1996) kaava.....                                    | 19 |
| 4.3.2 RATF .....  | 20 |
| 4.4 Palautteeseen valittavien laajennustermien määrä.....         | 22 |
| 4.5 Palautteeseen valittavien dokumenttien määrä .....            | 24 |
| <b>5. Palautekyselyn muodostaminen</b> .....                      | 27 |
| <b>6. Relevanssipalautteen tehokkuuden mittaaminen</b> .....      | 29 |
| <b>7. Yhteenvetoa teoriaosuudesta</b> .....                       | 31 |
| <b>8. Tutkimuskysymykset ja koeasetelma</b> .....                 | 35 |
| 8.1 Tutkimuskysymykset .....                                      | 35 |
| 8.2 Testikokoelma.....  | 36 |
| 8.3 Hakujärjestelmä .....   | 37 |
| 8.4 Kyselyjen suunnittelu.....                                    | 39 |
| 8.5 Tulosten analysointi ja evaluointi .....                      | 42 |
| 8.5.1 Hakutehokkuuden evaluointi .....                            | 42 |
| 8.5.2 Tilastolliset menetelmät .....                              | 43 |
| 8.6 Esitutkimus .....   | 44 |
| 8.7 Lopullinen koeasetelma .....                                  | 45 |
| <b>9. Tulokset</b> .....  | 47 |
| 9.1 Tulokset liberaaleilla relevanssikriteereillä mitattuna ..... | 48 |

|  |           |
|--|-----------|
| 9.2 Tulokset keskitiukoilla relevanssikriteereillä mitattuna .....                               | 50        |
| 9.3 Tulokset tiukoilla relevanssikriteereillä mitattuna.....                                     | 52        |
| 9.4 Yhteenvetoa tuloksista .....   | 54        |
| <b>10. Keskustelua</b> .....   | <b>58</b> |
| <b>11. Lähteet</b> .....   | <b>60</b> |
| <b>Liite</b> .....   | <b>64</b> |
| Merkitsevyystestien tulokset.....  | 64        |
| Laajennusmenetelmien vertailu, tuloksellisuus laskettu liberaaleilla relevanssikriteereillä..... | 65        |
| Laajennusmenetelmien vertailu, tuloksellisuus laskettu keskitiukoilla relevanssikriteereillä...  | 66        |
| Laajennusmenetelmien vertailu, tuloksellisuus laskettu tiukoilla relevanssikriteereillä .....    | 67        |

# 1. Johdanto

Tutkimuksessa tarkastellaan aloituskyselyn onnistuneisuuden vaikutusta relevanssipalautteen ja pseudorelevanssipalautteen tuloksellisuuteen. Relevanssipalautteella tarkoitetaan sitä, että käyttäjä valitsee tuloslistasta relevantiksi arvioimansa dokumentin tai useita dokumentteja ja hakujärjestelmä muodostaa dokumenteissa esiintyviä sanoja hyödyntäen muokatun kyselyn. Pseudorelevanssipalautteessa kone poimii automaattisesti tuloslistan kärjestä halutun määrän dokumentteja, joista se edelleen valitsee hakutermejä ja tekee näillä uuden kyselyn. Tämä tapahtuu piilossa loppukäyttäjältä ja vain lopullinen tuloslista tulee käyttäjälle näkyviin.

Relevanssipalautteessa ja pseudorelevanssipalautteessa on kysymys kyselyn laajentamisesta, jossa hyödynnetään käyttäjän relevanteiksi arvioimia dokumentteja. Hakujärjestelmä voi esimerkiksi lisätä dokumenttien termejä kyselyyn tai painottaa alkuperäisen kyselyn termejä uudestaan. Kyselyn laajennuksen taustalla on ajatus siitä, että käyttäjän tekemät kyselyt ovat usein lyhyitä, minkä seurauksena hakuaihetta kuvaavat vaihtoehtoiset ilmaiset ja näkökulmat eivät tule huomioiduksi. Laajennuksen avulla kyselystä saadaan kattavampi, jolloin on mahdollista löytää enemmän relevanteja dokumentteja.

Relevanssipalautetta on tutkittu paljon ja sen on varsin kiistattomasti todettu parantavan haun tuloksellisuutta. Relevanssipalautteen ja erityisesti pseudorelevanssipalautteen onnistuminen riippuu kuitenkin käyttäjän tekemästä kyselystä, sillä relevanssipalautte perustuu käyttäjän saamiin hakutuloksiin. Jos käyttäjän kysely epäonnistuu tai tuottaa vain vähän relevanteja dokumentteja, niin onko relevanssipalautteen avulla mahdollista paikata alkuperäisen kyselyn epäonnistuminen? Kuinka hyvin kyselyn oikeastaan pitäisi onnistua, jotta relevanssipalautteesta ja pseudorelevanssipalautteesta olisi apua? Erityisesti pseudorelevanssipalautteen osalta kysymys on kiinnostava. Vaikka pseudorelevanssipalautteella on onnistuttu parantamaan hakutehokkuutta, sen on myös huomattu heikentävän heikkotasoisten kyselyiden tuloksellisuutta.

Lähtötilanteena tutkimuksessa on kolmella eri tasolla onnistuneet, enintään kolmen sanan mittaiset kyselyt, joita kutsutaan tässä heikosti, hieman paremmin ja hyvin onnistuneiksi aloituskyselyiksi. Eritasoiset kyselyt olen muokannut hakuaiheiden otsikoissa ja relevanteissa dokumenteissa esiintyvien termien avulla. Oleellista tässä on, että kyselyt ovat eritasoisesti onnistuneita, ei se,

ovatko kyselyt hyviä tai huonoja sinänsä. Kokeellisissa tiedonhaku tutkimuksissa kyselyt muodostetaan tavallisesti automaattisesti hakuaiheiden otsikoiden tai kuvailukentissä esiintyvien sanojen perusteella. Tällöin kyselyistä tulee melko pitkiä, vaikka käyttäjien tekemät kyselyt ovat tunnetusti vain muutaman sanan mittaisia. Lisäksi automaattisesti muodostetuissa kyselyissä ei voida kontrolloida kyselyn onnistuneisuutta, vaikka käyttäjien tekemissä kyselyissä samasta hakuaiheesta esiintyy variaatiota onnistuneisuuden suhteen. Eritasoisesti onnistuneiden lyhyiden kyselyiden avulla voidaan tuoda kokeelliseen tutkimukseen realismia, sillä myös käyttäjien tekemät kyselyt onnistuvat vaihtelevasti ja ovat tavallisesti hyvin lyhyitä. Myös relevanssipalautteen määrä on rajattu kymmeneen dokumenttiin, sillä käyttäjä tuskin malttaa selata tuloslistaa kovin pitkälle.

Relevanssipalautteen toimivuutta eritasoisesti onnistuneilla aloituskyselyillä on tutkittu aika vähän. Käytännössä kyse on osittain siitä, kuinka monta relevanttia dokumenttia palautteeseen pitäisi valita optimaalisen hakutuloksen saavuttamiseksi, mikä puolestaan on huomattavasti tutkitumpi kysymyksen asettelu. Esimerkiksi Keskustalon, Järvelinin ja Pirkolan (2006) tutkimuksessa kävi ilmi, että relevanssipalautteeseen valittavien dokumenttien määrän kasvattaminen parantaa hakutehokkuutta. Pelkkä palautedokumenttien määrä ei kuitenkaan kerro suoraan kyselyn onnistuneisuudesta, sillä relevanttien dokumenttien lukumäärä hakuaiheisiin vaihtelee. Eritasoiset kyselyt sen sijaan voidaan suhteuttaa kyselykohtaisesti relevanttien dokumenttien määrää vastaaviksi.

Päätutkimuskysymykseni ovat:

## 1. Käyttäjän relevanssipalautteen tuloksellisuus

1.1 Miten aloituskyselyn onnistuneisuus vaikuttaa relevanssipalautteen tuloksellisuuteen?

1.2. Kuinka monta dokumenttia relevanssipalautteeseen pitäisi valita parhaan hakutuloksen saavuttamiseksi?

1.3 Minkä tasoisesti relevantteja dokumentteja relevanssipalautteeseen kannattaa valita, jotta palautteesta saadaan maksimaalinen hyöty?

## 2. Pseudorelevanssipalautteen tuloksellisuus

2.1 Miten aloituskyselyn onnistuneisuus vaikuttaa pseudorelevanssipalautteen tuloksellisuuteen?

2.2 Riippuuko pseudorelevanssipalautteen toimivuus palautteeseen valittujen dokumenttien lukumäärästä?

Tutkimus alkaa teoriaosuudella. Tämä jakautuu siten, että luvussa kaksi paikannan tutkimuksen osana tiedonhaketutkimuksen kenttää. Sitten hahmotan relevanssipalautetta suhteessa kyselyn laajentamiseen Efthimiadiksen (1996) esittämän mallin pohjalta. Tämän jälkeen esittelen pseudorelevanssipalautetta. Neljännessä luvussa esittelen, miten relevanssipalautetta on toteutettu vektorimalliin ja todennäköisyysmalliin pohjautuvissa tiedonhakujärjestelmissä. Lisäksi käsittelen palautteeseen valittavien laajennustermien valintamenetelmiä, joista yhtenä keskeisenä menetelmänä tarkastelen Pirkolan, Leppäsen ja Järvelinin (2002) kehittämää RATF:a, jota käytetään myös tässä tutkimuksessa. Luvussa 5 käsittelen lyhyesti palautekyselyn muodostamista. Lopuksi tarkastelen erilaisia menetelmiä, joiden avulla voidaan evaluoida relevanssipalautehaun tuloksellisuutta.

## 2. Tiedonhaun tutkimus

Ennen kuin päästään käsiksi kyselyn laajentamiseen ja relevanssipalautteeseen, on syytä kertoa hieman, mistä tiedonhaussa on kysymys. Tässä luvussa esittelen tiedonhaun tutkimukseen liittyviä peruskäsitteitä, kuten tiedonhakua, täsmäytysmenetelmiä ja relevanssia.

### 2.1 Tiedonhaku ja tiedonhakuprosessit

Sanalla tiedonhaku voidaan viitata useaan eri asiaan. Se voi tarkoittaa mm. tutkimusalaa, hakutehtävää, hakutulosta tai hakutehtävään perustuvaa kyselyä. (Järvelin 1995, 55.) Tiedonhaussa on kysymys dokumenttien organisoinnista, säilyttämisestä ja esittämisestä. Dokumenttien esittämisen ja organisoimisen pitäisi taata käyttäjälle pääsy tiedonlähteisiin. Näistä nousevia ongelmia tiedonhaketutkimus pyrkii ratkomaan. (ks. esim. Baeza-Yates ja Ribeiro-Neto 1999, 1 ja Kekäläinen 1999, 1) Tässä tutkimuksessa tiedonhaulla tarkoitetaan hakutehtävään perustuvaa kyselyä. Tiedonhakua ei kuitenkaan tarvittaisi ilman tiedontarpeita. Tiedontarpeella tarkoitetaan ongelmaa, jonka ratkaisemiseksi tiedontarvitsijan on ryhdyttävä toimenpiteisiin. Tällainen toimenpiteeksi aktualisoitunut tiedontarve on esimerkiksi käyttäjän muodostama kysely tiedonhakuprosessiin. Tiedonhakuprosessi on laajasti määriteltynä tietoyksiköiden, kuten esimerkiksi tekstidokumenttien tallentamiseen, jakeluun ja hakuun käytettävä järjestelmä. Tällaisia ovat esimerkiksi erilaiset rekisterit ja kirjastojen viitetietokannat. (Järvelin 1995, 20-21.)

Tiedonhaun tutkimuksen yksi perusongelma on se, että tiedontarvetta ei voida täsmällisesti kuvata hakutehtävänä, ennen kuin vastaus on löydetty. Mikäli ”täydellinen” haku pystyttäisiin muotoilemaan, olisi vastauksenkin jo oltava tiedossa. (ks. esim. Järvelin 1995, 28) Toinen ongelma liittyy tiedontarpeen dynaamisuuteen. Tiedontarpeet muuttuvat hakuprosessin edetessä, jolloin myös käsitykset hyödyllisistä dokumenteista muuttuvat.

Perinteisissä tiedonhakuprosesseissa dokumenttikanta on suhteellisen muuttumaton, mutta järjestelmään tehdyt kyselyt sen sijaan vaihtelevat. Tällaista hakuoperaatiota kutsutaan ad hoc -hauksi ja se on hyvin yleinen hakutyyppi. Toisen tyyppinen hakuoperaatio on suodatus (filtering), jossa hakutehtävät pysyvät suhteellisen muuttumattomina, mutta dokumenttikokoelma sen sijaan täydentyy uusilla dokumenteilla. (Baeza-Yates ja Ribeiro-Neto 1999, 21-22.) Tässä tutkimuksessa operoidaan ad hoc -hauilla. Tarkoituksena on kokeellisesti muokata erityyppisiä kyselyitä, joita ajetaan dokumenttikannassa.



## **2.2 Täsmäytysmenetelmät**

Jotta käyttäjän tekemällä kyselyllä olisi ylipäättään mahdollista löytää dokumentteja tietokannasta, tarvitaan täsmäytysmenetelmää. Täsmäytysmenetelmällä tarkoitetaan hakujärjestelmän käyttämää algoritmia, joka vertailee kyselyä ja tietokannan dokumentteja toisiinsa ja palauttaa käyttäjälle tulosjoukon. Tulosjoukon dokumenttien lajittelu riippuu käytetystä täsmäytysmenetelmästä. Täsmäytysmenetelmät voidaan jakaa karkeasti kahteen pääkategoriaan: täystäsmäyttäviin ja osittaistäsmäyttäviin. Täystäsmäyttävä haku perustuu tavallisesti Boolean logiikkaan, kun taas yleisiä osittaistäsmäyttäviä menetelmiä ovat esimerkiksi vektorimalli ja todennäköisyyksiin perustuva probabilistinen malli. (ks. esim. Kekäläinen 1999, 16.)

Täystäsmäyttävä järjestelmä perustuu siihen, että kyselyn on vastattava täysin tietokannassa olevaa dokumenttia, jotta dokumentti täsmäytyy tulosjoukkoon. Menetelmän ongelmana on se, että dokumentit, jotka sisältävät vain osan kyselyssä esiintyvistä sanoista, eivät päädy tulosjoukkoon. Se ei myöskään osaa lajitella tulosjoukkoa relevanssin mukaan ja vaatii lisäksi käyttäjältä tietämystä Boolean logiikasta. Osittaistäsmäyttävä hakujärjestelmä korjaa monia täystäsmäyttävän järjestelmän puutteita. Ensinnäkään dokumentin ei tarvitse sisältää kaikkia kyselyn termejä päätyäkseen tulosjoukkoon. Toiseksi se mahdollistaa relevanssilajittelun. (ks. esim. Kekäläinen 1999, 16.) Täsmäytysmenetelmiin palataan vielä luvuissa 4.1 ja 4.2 relevanssipalautteen yhteydessä.

## **2.3 Relevanssin käsitteestä**

Hakujärjestelmän palauttamia dokumentteja käyttäjän tekemään kyselyyn arvioidaan dokumenttien relevanssin perusteella. Relevanssi on informaatiotutkimuksessa perinteisesti jaettu kahteen osaluokkaan: käyttäjärelevanssiin ja aiherelevanssiin. Ensiksi mainitulla tarkoitetaan sitä, että dokumentti on relevantti, mikäli se tyydyttää käyttäjän tiedontarpeen. Viimeksi mainitulla tarkoitetaan puolestaan sitä, että dokumentti on relevantti, mikäli se vastaa esimerkiksi indeksointitermeillä kuvailtua dokumenttia. Pelkistetysti aiherelevanssi viittaa kyselyn ja dokumentin sanojen täsmäytykseen. Relevanssi voidaan nähdä myös käyttäjän tulkitseman tiedon ja tiedontarvetilanteen välisenä suhteena, jolloin relevanssi muuttuu tilannesidonnaiseksi ja dynaamiseksi käsitteeksi. (Järvelin 1995, 46.)

Aiherelevanssi on tiedonhaku tutkimuksessa käyttökelpoinen sen mitattavuuden vuoksi. Se on riittävä relevanssin määritelmä myös tämän tutkimuksen tarpeisiin. Oletuksena on, että hyvä hakuavainten täsmäytys johtaa hyvään tuloksellisuuteen myös tiedonhaussa (Järvelin 1995, 46-47).

## **2.4 Relevanssiarvioinneista tiedonhaun tutkimuksessa**

Yksi keskeinen ongelma tiedonhaun tutkimuksessa liittyy relevanssiarvioon. Monissa laboratoriotutkimuksissa on vielä viimevuosiin saakka ollut käytössä kaksiportainen relevanssiasteikko, jossa dokumentit arvioidaan joko relevanteiksi tai epärelevanteiksi. Lisäksi relevanssikriteerit ovat olleet melko liberaaleja, jolloin relevantiksi arvioitu dokumentti saattaa liittyä vain kevyesti hakuaiheeseen. (Sormunen 2002a.) Tässä ei tule huomioiduksi se, että jotkut dokumentit voivat olla enemmän tai vähemmän relevantteja kuin toiset. Tämä on tärkeä huomio, sillä tietokantojen paisuessa yhä suuremmiksi, saattaa relevantteja dokumentteja löytyä huomattavasti enemmän, kuin mitä käyttäjä on halukas käymään läpi. Näin ollen käyttäjän näkökulmasta olisi toivottavaa, että tuloslistan kärkeen osuu kaikkein relevantteimmat dokumentit. (Kekäläinen ja Järvelin 2002.)

Tässä tutkimuksessa hyödynnetään neliportaista relevanssiasteikkoa, jossa dokumentti voi olla relevantin lisäksi erittäin relevantti, marginaalisesti relevantti tai epärelevantti. Tähän on päädytty osittain käytännön syistä, sillä tutkimuksen tarpeisiin sopiva dokumenttikokoelma relevanssiarvioineen on valmiiksi olemassa.

### **3. Kyselyn laajentaminen**

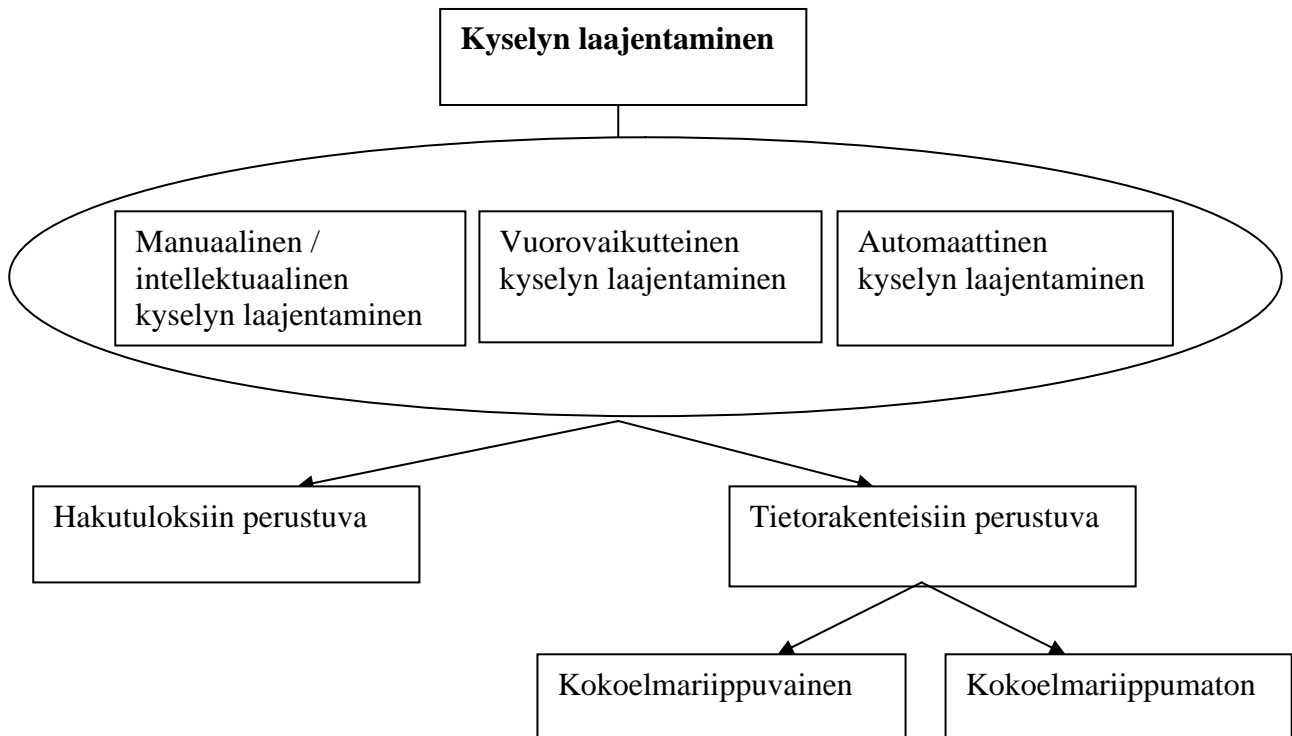
Relevanssipalautteessa on kysymys kyselyn laajentamisesta, jossa hyödynnetään käyttäjän relevanteiksi arvioimia dokumentteja. Hakujärjestelmä voi esimerkiksi lisätä dokumenttien termejä kyselyyn tai painottaa alkuperäisen kyselyn termejä uudestaan. Tässä luvussa käsitellään aluksi laajemmin kyselyn laajentamista Efthimiadiksen (1996, 122) esittämän mallin pohjalta. Tämän jälkeen perehdytään tarkemmin relevanssipalautteeseen ja pseudorelevanssipalautteeseen.

#### ***3.1 Kyselyn laajentamisen malli***

Kyselyn muotoilua, uudelleenmuotoilua ja laajennusta on tutkittu paljon, sillä hyvien hakutermien valinta on tärkeää hyvän hakutuloksen saavuttamiseksi. Käyttäjien tekemät kyselyt eivät useinkaan sisällä kaikkia hakuaiheen kannalta tarpeellisia termejä, eivätkä kyselyissä käytetyt termit ole välttämättä hakuaiheen kannalta parhaita. Tavallisesti kyselyt ovat lyhyitä ja hakuaiheen eri näkökulmia kuvataan korkeintaan yhdellä hakutermillä. (ks. esim. Kekäläinen 1999, 32.) Yksi ratkaisu kyselyn muotoilun aiheuttamiin ongelmiin on kyselyn laajentaminen. Kyselyn laajentamisella tarkoitetaan uusien hakuavainten lisäämistä käyttäjän tekemään kyselyyn tai hakuavainten painottamista uudelleen, käytännössä usein molempia.

Efthimiadis (1996, 122-124) jakaa kyselyn laajentamisen ensinnäkin kolmeen eri kategoriaan sen mukaan, miten laajennustermit valitaan. Näitä kategorioita ovat intellektuaalinen (tai manuaalinen), automaattinen ja vuorovaikutteinen kyselyn laajentaminen. Intellektuaalisessa kyselyn laajentamisessa käyttäjä valitsee itse termit laajennukseen. Tämä vaatii käyttäjältä hakustrategiaa, eli suunnitelmaa haun toteutuksesta, asiantuntemusta tiedonhakujärjestelmän toiminnallisuudesta ja hakuaiheen tuntemista. (Efthimiadis 1996, 126.) Vuorovaikutteisessa kyselyn laajentamisessa hakujärjestelmä ehdottaa käyttäjälle mahdollisia hakutermejä, joista käyttäjä valitsee haluamansa termit laajennukseen. Vastuu vuorovaikutteisen kyselyn laajentamisen onnistumisesta on niin käyttäjällä kuin hakuavaimia ehdottavalla hakujärjestelmällä. (Efthimiadis 1996, 156-157.) Automaattisessa kyselyn laajentamisessa käyttäjä ei voi vaikuttaa laajennustermien valintaan. Tässä hakujärjestelmä poimii hakutermejä joko käyttäjän valitsemista relevanteista dokumenteista tai vaihtoehtoisesti hakujärjestelmä voi valita myös dokumentit, joiden perusteella laajennus tehdään.

Tällöin laajennuksen onnistuminen riippuu lähinnä käyttäjän tekemästä hausta ja algoritmista, jolla laajennustermit valitaan. (Efthimiadis 1996, 122.)



**Kuva 1. Kyselyn laajentamisen malli** (Efthimiadis 1996, 124)

Toisekseen Efthimiadis (1996, 122) jaottelee kyselyn laajentamisen kahteen osa-alueeseen sen mukaan mistä lähteistä laajennustermit valitaan. Laajennustermit voidaan valita ensinnäkin hakutulosten joukosta, kuten esimerkiksi relevanssipalautteessa tehdään. Tähän palataan tarkemmin seuraavassa luvussa. Toiseksi laajennustermit voidaan valita tietorakenteista. Tietorakenteet Efthimiadis jaottelee sen mukaan, ovatko ne sidoksissa dokumenttikokoelmaan vai kokoelmasta riippumattomia. Kokoelmasta riippuvainen tietorakenne voi olla vaikkapa automaattisesti muodostettu tesaurus. Kokoelmasta riippumaton tietorakenne on esimerkiksi intellektuaalisesti muodostettu tesaurus kuten vaikkapa lääketieteen asiasanasto MeSH.

## **3.2 Relevanssipalautte**

Relevanssipalautteella tarkoitetaan sitä, että käyttäjä tekee kyselyn, valitsee tuloslistasta relevantiksi arvioimansa dokumentin tai useita dokumentteja ja hakujärjestelmä muodostaa dokumenteissa esiintyvillä sanoilla uuden kyselyn. Relevanssipalautte on toisin sanoen yksi tapa laajentaa kyselyä. Relevanssipalautteen lähtökohtana on ajatus siitä, että tiedonhakijalla ei ole välttämättä kovin tarkkaa käsitystä siitä, minkälaista tietoa hän on hakemassa. Hakijat käyttävät yksinkertaisia kyselyitä ja erilaisten vaihtoehtoisten ilmausten huomiointi saattaa tuottaa vaikeuksia. Tutkijat ovat kuitenkin jo aikaisemmin huomanneet, että vaikka tiedontarpeen muotoileminen kyselyksi saattaa tuottaa käyttäjille ongelmia, he kuitenkin tunnistavat relevantin dokumentin sellaisen nähdessään. Tämä on johtanut ajatukseen relevanssipalautteesta. (Ruthven ja Lalmas 2003.)

Keskeinen kysymys relevanssipalautteessa liittyy relevanssitietojen keräämiseen, sillä puutteellisilla relevanssitiedoilla relevanssipalautteen hyöty voi jäädä vähäiseksi. Jimmy Lin (2006) jakaa relevanssipalautteen kolmeen osa-alueeseen sen mukaan, miten relevanssitietoja kerätään:

1. Eksplisiittisessä relevanssipalautteessa käyttäjä erottelee relevantit ja epärelevantit dokumentit.
2. Implisiittisessä relevanssipalautteessa järjestelmä havainnoi käyttäjän toimia ja tekee tämän perusteella päätelmiä relevanssista. Tämä voidaan toteuttaa esimerkiksi seuraamalla, kuinka pitkään käyttäjä tarkastelee tuloslistan dokumentteja. Mitä kauemmin käyttäjä viihtyy dokumentin parissa, sitä todennäköisemmin se on relevantti.
3. Sokeassa relevanssipalautteessa (pseudorelevanssipalautteessa) relevanssitietoja ei kerätä lainkaan. Voidaan esimerkiksi olettaa, että tulosjoukon kärjessä on relevantteja dokumentteja ja kerätä laajennustermejä näistä dokumenteista.

Eksplisiittisellä relevanssipalautteella voidaan laajentaa kyselyä ainakin kolmella eri menetelmällä sen mukaan, mistä laajennustermit valitaan, kuinka termejä painotetaan ja miten alkuperäisen kyselyn termejä käsitellään.

- Ensinnäkin relevanssipalautteessa voidaan hyödyntää pelkästään alkuperäisessä kyselyssä mukana olevia termejä painottamalla termejä uudelleen relevanteissa dokumenteissa esiintyvien termien perusteella. Termien painottamiseen palataan tarkemmin luvuissa 4.2 ja 4.3.
- Toiseksi laajennuksessa voidaan hyödyntää alkuperäisten termien ohella tuloslistan dokumenteissa esiintyviä termejä.

- Kolmanneksi laajennustermit voidaan valita pelkästään tulosjoukon dokumenteista ja hylätä alkuperäisen kyselyn termit kokonaan. (Efthimiadis 1996, 134-135.)

Loppukäyttäjän kannalta näissä menetelmissä on kysymys relevanttien dokumenttien valinnasta, joiden perusteella hakujärjestelmä muotoilee kyselyn uudelleen, eikä käyttäjällä ole mahdollisuutta puuttua varsinaisen laajennuksen muotoilemiseen. Tämän kaltainen automaattinen relevanssipalautte voidaan erottaa interaktiivisesta relevanssipalautteesta, jossa käyttäjä osallistuu myös varsinaiseen kyselymuotoiluprosessiin esimerkiksi valitsemalla hakutermejä tai painottamalla hakusanoja. Tällöin voidaan puhua myös ns. termirelevanssipalautteesta. Tässä työssä relevanssipalautteesta puhuttaessa viitataan eksplisiittiseen relevanssipalautteeseen, jossa käyttäjä valitsee palauttedokumentit, mutta ei osallistu varsinaisten laajennustermien valintaan.

### **3.3 Pseudorelevanssipalautte**

Pseudorelevanssipalautteen ajatuksena on toimia relevanssipalautteen tapaan, eli laajentaa kyselyä tulosjoukon dokumenttien perusteella. Pseudorelevanssipalautteessa yhtään dokumenttia ei kuitenkaan näytetä käyttäjälle ennen lopullista tuloslistaa, vaan hakujärjestelmä valitsee tulosjoukon kärjestä vakiomäärän dokumentteja olettaen, että ne ovat kaikki keskimäärin relevantteja. Näistä sokkona valituista dokumenteista hakujärjestelmä valitsee termejä, lisää termit alkuperäiseen kyselyyn, muodostaa uuden kyselyn ja vasta tämän tulosjoukon loppukäyttäjää saa nähtäväkseen. (ks. esim. Ruthven ja Lalmas 2003, 31.)

Käyttäjän näkökulmasta pseudorelevanssipalautteessa kyse ei oikeastaan ole relevanssipalautteesta, sillä käyttäjälle prosessi ei ilmene millään tavalla, eikä palautetta anneta missään vaiheessa. Teknisessä mielessä pseudorelevanssipalautte voidaan kuitenkin toteuttaa samaan tapaan kuin aito relevanssipalautte. Erottavana tekijänä on lähinnä mekanismi, jolla dokumentit valitaan palautteeseen.

Pseudorelevanssipalautteen voidaan olettaa parantavan hyvin onnistuneiden kyselyiden tuloksellisuutta ja heikentävän alun perin heikosti onnistuneiden kyselyiden tuloksellisuutta entisestään. Jos tulosjoukossa on pääosin epärelevantteja dokumentteja, on todennäköistä, että näistä dokumenteista poimituilla termeillä löydetään lisää epärelevantteja dokumentteja. Mikäli

tulosjoukon dokumenteista valtaosa on epärelevantteja, on vaarana, että näistä dokumenteista valikoituu termejä myös laajennukseen ja tämän seurauksena laajennus heikentää hakutehokkuutta. Ongelmaan voidaan tarjota lääkkeeksi joitakin menetelmiä. Ensinnäkin voidaan yrittää muokata alkuperäistä tuloslistaa ja kasvattaa todennäköisyyttä sille, että valtaosa palautteeseen valituista dokumenteista on relevantteja. Toinen vaihtoehto on kehittää termiervalintamenetelmiä, jotka osaavat jättää huonot hakutermit pois laajennuksesta. (Ruthven ja Lalmas 2003, 30.)

Kolmanneksi järjestelmä voisi yrittää päätellä käyttäjän tekemästä kyselystä ja kyselyn tuottamasta tulosjoukosta, onko kysely onnistunut. Mikäli järjestelmä arvioi kyselyn pieleen menemisen todennäköisyyden tarpeeksi suureksi, pseudorelevanssipalautetta ei tehtäisi lainkaan. Kyselyn epäonnistumista voitaisiin ennakoida esimerkiksi kyselyn pituudella. Jos kyselyssä on esimerkiksi vain yksi tai kaksi hakusanaa, on kohtuullisen todennäköistä, että relevantteja dokumentteja ei löydy kovin paljon. Toinen tapa päätellä kyselyn onnistuneisuutta voisi olla tuloslistan dokumenttien vertaileminen. Mikäli dokumenteissa on hyvin vähän yhteisiä piirteitä, voidaan olettaa, että kysely on hajaantunut liikaa, eikä pseudopalautetta kannata tehdä.

## 4. Relevanssipalautetekniikoita

Tässä luvussa esittelen tarkemmin, miten relevanssipalautte voidaan toteuttaa vektorimalliin ja todennäköisyysmalliin perustuvissa tiedonhakujärjestelmissä. Lisäksi esittelen todennäköisyysmallissa toteutettuja menetelmiä kyselytermien painotukselle.

### 4.1 Relevanssipalautte vektorimallissa

Tiedonhakujärjestelmien relevanssipalautte esiteltiin ensimmäisen kerran jo 60-luvun lopulla, jonka jälkeen sitä on pyritty tehostamaan monenlaisilla tekniikoilla (Lundquist, Grossman ja Frieder 1997). Yksi varhaisimmista tutkimuksista on Rocchion (1971) tekemät kokeet vektorimalliin perustuvalla SMART -tiedonhakujärjestelmällä. Hänen mukaansa tiedonhaussa on kysymys optimaalisen kyselyn muodostamisesta. Optimaalinen kysely maksimoi relevanttien dokumenttien muodostaman keskimääräisen vektorin ja epärelevanttien dokumenttien muodostaman keskimääräisen vektorin välisen eron. Käyttäjän ei kuitenkaan ole useinkaan mahdollista tehdä tällaista ”optimaalista” kyselyä, jolloin tarvitaan relevanssipalautetta siirtämään käyttäjän kyselyvektori lähemmäksi relevanttien kyselyiden vektoria. Tämä voidaan saavuttaa lisäämällä käyttäjän tekemään kyselyyn uusia termejä ja painottamalla termejä sen mukaan, kuinka hyviä ne ovat erottelemaan relevantteja dokumentteja epärelevanteista. Uusi kyselyvektori on näin ollen alkuperäinen kyselyvektori, johon on lisätty relevanttien dokumenttien termeistä muodostettu kyselyvektori. (Ruthven ja Lalmas 2003.) Seuraavaksi esitellään tarkemmin relevanssipalautteen toiminta vektorimallissa Saltonin ja Buckleyyn (1990) artikkelin pohjalta.

Vektoriympäristössä dokumentit  $D$  ja kyselyt  $Q$  voidaan esittää  $t$ -ulotteisina vektoreina muodossa  $D = (d_1, d_2, \dots, d_t)$  ja  $Q = (q_1, q_2, \dots, q_t)$ . Tällöin  $d_i$  ja  $q_i$  ovat painoja termille  $i$  dokumentissa  $D$  ja kyselyssä  $Q$ . Oletetaan lisäksi, että vektoreiden yhteen- ja vähennyslasku sekä kerto- ja jakolasku on määritelty. Kyselyä ja dokumenttia vastaava samankaltaisuus voidaan tällöin laskea niitä vastaavien vektoreiden sisätulona, joka voidaan esittää muodossa

(Kaava 1)

$$Sim(D, Q) = \sum_{i=1}^t d_i \cdot q_i$$



Näin saadaan lasketuksi vektoreiden  $D$  ja  $Q$  välinen kulma. Kun hakujärjestelmässä käytetään dokumenttivektorin ja kyselyvektorin samankaltaisuuteen perustuvia laskelmia, voidaan optimaalinen kyselyvektori esittää seuraavasti:

$$(Kaava 2) \quad Q_{opt} = \frac{1}{n} \sum_{\text{relevantit}} \frac{D_i}{|D_i|} - \frac{1}{N-n} \sum_{\text{epärelevantit}} \frac{D_i}{|D_i|}$$

$D_i$  viittaa tässä dokumenttivektoriin ja  $|D_i|$  on vastaavan euklidisen vektorin pituus.  $N$  on dokumenttikannan koko ja  $n$  on relevanttien dokumenttien määrä kokoelmassa. Kaavaa 2 ei kuitenkaan voida käyttää todellisuudessa, sillä relevanttien dokumenttien määrä ei ole etukäteen tiedossa. Kaavaa voidaan kuitenkin hyödyntää palautteessa, kunhan relevanssitetiedot löydettyistä dokumenteista on tiedossa. Tällöin relevanttien ja epärelevanttien dokumenttien paikalle sijoitetaan *tunnetut* relevantit ja *tunnetut* epärelevantit dokumentit. Aikaisempien tutkimusten perusteella lisäksi tiedetään, että alkuperäisen kyselyn termit kannattaa sisällyttää hakuun. Palautekysely voidaan tällöin esittää, seuraavalla kaavalla, jossa  $n_1$  tarkoittaa relevanttien dokumenttien lukumäärää ja  $n_2$  epärelevanttien dokumenttien lukumäärää,  $Q_0$  alkuperäistä kyselyä ja  $Q_1$  ensimmäisen palautekierroksen jälkeen muodostettua kyselyä.

$$(Kaava 3) \quad Q_1 = Q_0 + \frac{1}{n_1} \sum_{\text{tunnetut relevantit}} \frac{D_i}{|D_i|} - \frac{1}{n_2} \sum_{\text{tunnetut epärelevantit}} \frac{D_i}{|D_i|}$$

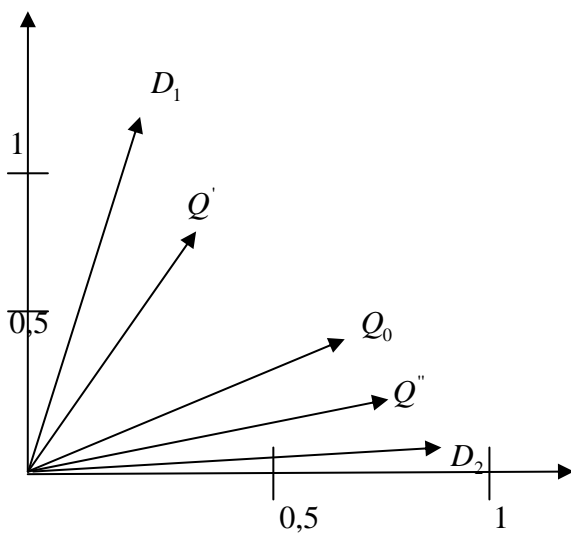
Tämän jälkeen lasketaan relevanttien ja epärelevanttien dokumenttien välinen summa. Kyselyn muodostamisessa voidaan lisäksi hyödyntää painokertoimia  $\alpha$ ,  $\beta$  ja  $\gamma$ , jotka voidaan normalisoida vaihteluvälille 0 ja 1 (Kaava 4).

$$(Kaava 4) \quad Q_{i+1} = \alpha Q_i + \beta \sum_{\text{tunnetut relevantit}} \frac{D_i}{|D_i|} - \gamma \sum_{\text{tunnetut epärelevantit}} \frac{D_i}{|D_i|}$$

Relevanssipalaute on havainnollistettu kuvassa 2, joka mallintaa kaksiulotteista vektoriavaruutta ja jossa jokaisella vektorilla on kaksi komponenttia. Oletetaan, että dokumentti  $D_1$  on määritelty relevantiksi alkuperäiseen kyselyyn  $Q_0$ . Kun uusi kysely muodostetaan dokumentin  $D_1$  avulla,

saadaan uusi kysely  $Q'$ , joka on huomattavasti alkuperäistä kyselyä lähempänä dokumenttia  $D_1$ . Uudella kyselyllä  $Q'$  voidaan olettaa löytyvän lisää dokumentin  $D_1$  kaltaisia relevantteja dokumentteja. Vastaavalla tavalla käy, kun dokumentti  $D_2$  tunnistetaan relevantiksi alkuperäiseen kyselyyn ja lasketaan tämän perusteella uusi kysely  $Q''$ .

**Kuva 2. Relevanssipalaute kaksiulotteisessa vektoriavaruudessa**



#### **4.2 Relevanssipalaute todennäköisyysmallissa**

Todennäköisyysmalli perustuu sen arvioimiseen, millä todennäköisyydellä dokumentti on relevantti käyttäjän tekemään kyselyyn. Mitä korkeammaksi todennäköisyys on arvioitu, sitä todennäköisemmin dokumentti on relevantti. Näin kyselyyn tulokseksi saadut dokumentit voidaan esittää relevanssilajiteltuna tuloslistana lasketun todennäköisyyden mukaan.

Tässä esitetty todennäköisyysmalli perustuu binääriseen riippumattomuusmalliin, joka on perinteisin todennäköisyyksiin perustuva lähestymistapa. Todennäköisyysmallissa oletuksena on, että termit esiintyvät toisistaan riippumattomina. Tämä tarkoittaa, että jos dokumentissa esiintyy termi  $t$ , se ei vaikuta termin  $s$  esiintymiseen samassa dokumentissa. Riippumattomuusoletusta voidaan tarkentaa seuraavasti:

11. Termien jakautuminen relevanteissa dokumenteissa on riippumatonta ja termien jakautuminen kaikissa dokumenteissa on riippumatonta.

12. Termien jakautuminen relevanteissa dokumenteissa on riippumatonta ja termien jakautuminen epärelevanteissa dokumenteissa on riippumatonta.

Toisin sanoen kysymys on siitä, verrataanko termin esiintymisen todennäköisyyttä suhteessa epärelevantteihin dokumentteihin vai kaikkiin kokoelman dokumentteihin. Tätä voidaan havainnollistaa kuvalla 3

|           |       |           |       |
|-----------|-------|-----------|-------|
|           | Rel   | Erel      |       |
| $x_t = 1$ | $r$   | $n-r$     | $n$   |
| $x_t = 0$ | $R-r$ | $N-n-R+r$ | $N-n$ |
|           | $R$   | $N-R$     |       |

**Kuva 3. Taulukko termien esiintymisestä**

$r$  = relevanttien dokumenttien määrä, jotka sisältävät termin  $x_t$

$n$  = dokumenttien lukumäärä, jotka sisältävät termin  $x_t$

$R$  = relevanttien dokumenttien lukumäärä kyselyyn  $q$

$N$  = dokumenttien lukumäärä kokoelmassa

Todennäköisyysmalli listaa dokumentit relevanssin mukaiseen järjestykseen sen mukaan, kuinka todennäköisesti dokumentit ovat relevantteja kyselyyn. Tästä ns. lajitteluperiaatteesta voidaan erottaa kaksi versiota sen mukaan a) perustuuko todennäköisyys kyselytermin esiintymiselle dokumentissa vai b) otetaanko tämän lisäksi huomioon kyselytermien puuttuminen dokumenteista. Tässä termien painotuksissa käytetään kaavaa, jossa lähtökohtana on riippumattomuusoletus 12 ja lajitteluperiaate b). Se voidaan esittää seuraavasti:

(Kaava 5) 
$$w_t = \log \frac{p_t(1 - q_t)}{q_t(1 - p_t)}$$

$w_t$  = termin  $t$  paino

$p_t$  = todennäköisyys sille, että termi  $t$  esiintyy relevantissa dokumentissa. Tämän laskemiseen voidaan soveltaa termien frekvenssejä  $r/R$ , jotka saadaan kuvan 3 taulukosta.

$q_i$  = todennäköisyys sille, että termi  $t$  esiintyy epärelevantissa dokumentissa. Tämän laskemiseen voidaan soveltaa termien frekvenssejä  $(n-r)/(N-R)$  (ks. kuva 3).

Kaavan 5 hyödyntäminen edellyttää tietoja termin esiintymisistä relevanteissa ja epärelevantissa dokumenteissa.  $p:n$  ja  $q:n$  arviointiin voidaan kuitenkin käyttää relevanssipalautteesta saatuja tietoja, jolloin kaava voidaan laskea esimerkiksi tuloslistan kymmenen ensimmäisen dokumentin joukosta. Käytännössä todennäköisyydet voidaan korvata termien frekvensseillä (ks. kuva 3), jolloin painotuskaava saadaan seuraavaan muotoon.

$$(Kaava 6) \quad f4 = w_t = \log \frac{r(N - n - R + r)}{(n - r)(R - r)}$$

Kirjallisuudessa kaava 6 tunnetaan  $f4$  -painotusfunktiona, joka on muodostunut jonkinasteiseksi vertailufunktioksi, kun on tutkittu erilaisten painotuskaavojen tehokkuuksia (ks. esim. Efthimiadis 1995).  $f4$  -painotusfunktio laskee painoja termeille sen mukaan, miten termit jakaantuvat relevanttien ja epärelevanttien dokumenttien kesken. Alkutilanteessa ei ole relevanssi-informaatiota tarjolla, minkä vuoksi todennäköisyyksiä ei voida arvioida. Tämä voidaan ratkaista käyttämällä vaikkapa  $idf$  -painotusfunktiota (esitellään tarkemmin luvussa 4.3). Kun ensimmäinen tuloslistaus on saatu, voidaan näistä dokumenteista antaa relevanssitiedot ja hyödyntää kaavaa 6 termien uudelleenpainotuksessa. Toinen ratkaisu relevanssi-informaation puutteeseen on pseudorelevanssipalautteen käyttäminen. Tällöin oletetaan, että tietty määrä kärkidokumentteja on relevantteja ja lasketaan painot tämän ”relevanssitiedon” pohjalta. (Ruthven ja Lalmas 2003.)

Kaavan 6 ongelmaksi muodostuu tilanne, jossa  $r$  ja  $R$  saavat arvon nolla. Tähän on esitetty ratkaisuksi, että  $r$  korvataan lausekkeella  $(r+n/N)$  (tässä Efthimiadis 1995), mikä kuvan 3 pohjalta tarkoittaa termin  $t$  sisältävien dokumenttien suhdetta kaikkiin tietokannan dokumentteihin

$$(Kaava 7) \quad f4 \text{ mod} = \log \frac{(r + n/N)(N - n - R + r + n/N)}{(n - r + n/N)(R - r + n/N)}$$

$f4$  -kaavaa tai sen läheisiä variantteja voidaan käyttää dokumenttien täsmäytykseen siten, että lasketaan kyselytermien painojen summa dokumentissa. Termien valintaan relevanssipalautetta varten se ei välttämättä ole tarkoituksenmukaisin. Se, että termi toimii dokumentin relevanssin indikaattorina, ei välttämättä tarkoita sitä, että samalla termillä voidaan parantaa hakutehokkuutta,

mikäli se lisätään kyselyyn. Sen sijaan kaavalla 8 voidaan lajitella laajennustermejä sen mukaan, miten termi jakautuu relevanttien ja epärelevanttien dokumenttien kesken.  $a_t$  :n mukaan lajitellusta listasta voidaan valita haluttu määrä termejä laajennukseen. (Ks. myös Robertson 1990.)

(Kaava 8)

$$a_t = w_t(p_t - q_t)$$

$a_t$  = termin  $t$  arvo laajennusterminä

$w_t$  = termin  $t$  paino

Efthimiadis (1995) on testannut kaavan 8 pohjalta  $wpq$  -painotuskaavaa, jossa  $w$  on korvattu aiemmin esitetyllä  $f4$  -funktiolla. Tällöin saadaan kuvan 3 tietoja hyväksi käyttäen kaava 9, jossa  $r$  on korvattu lauseella  $r + 0,5$  nolatilanteiden välttämiseksi.

(Kaava 9)

$$wpq = \log \frac{(r + 0,5)(N - n - R + r + 0,5)}{(n - r + 0,5)(R - r + 0,5)} * \left( \frac{r}{R} - \frac{n - r}{N - R} \right)$$

Minkälaiset laajennustermit sitten ovat hyödyllisiä eli hyviä erottelemaan epärelevantteja dokumentteja relevanteista? Efthimiadiksen (1995) päättelee aiemman tutkimuksen perusteella seuraavaa:

- termit, jotka esiintyvät usein dokumentissa , eivät ole laajennuksen kannalta hyödyllisiä
- kohtuullisen usein esiintyvät termit (middle-frequency terms) ovat melko hyödyllisiä
- harvinaiset termit ovat todennäköisesti hyödyllisiä, mutta eivät niin hyviä kuin kohtuullisen usein esiintyvät termit
- erittäin harvinaiset termit ovat hyviä indikoimaan relevanssia silloin, kun ne esiintyvät dokumentissa

Tämän pohjalta voidaan sanoa, että hyödyllinen algoritmi poimii tehokkaasti termifrekvenssiltään keskimääräisiä termejä laajennustermilistan kärkeen.

Efthimiadis (1995) vertaili edellä esiteltujen painotusfunktioiden  $wpq$ ,  $f4_{mod}$  ja  $f4$  (0,5 painotuksella) toimivuutta vuorovaikutteisessa relevanssipalautteessa (termirelevanssipalautteessa). Mukana vertailussa oli muitakin funktioita, joista mainittakoon tässä hänen itse kehittämänsä kaava *r-lohi*. Kaavassa on ajatuksena, että listataan relevanteissa dokumenteissa esiintyvät termit  $r$  frekvenssin mukaisessa laskevassa järjestyksessä ja valitaan laajennukseen termit, joilla on suurin

(high) frekvenssi. Mikäli termeillä on sama frekvenssi, suositaan sellaisia termejä, joilla on pienin (low) frekvenssi kaikissa dokumenteissa  $n$ . Kaavoja käytettiin tutkimuksessa termien poimintaan relevanteista dokumenteista ja myös termien painottamiseen. Koska kyseessä oli vuorovaikutteinen tutkimus, oli käyttäjälähtöisyys mukana myös kaavojen tehokkuuden arvioinnissa. Kaavojen tehokkuutta mitattiin siten, että katsottiin mitä termejä käyttäjät valitsevat laajennukseen ja katsottiin valitsevatko kaavat samoja laajennustermejä kuin käyttäjät. Ajatuksena oli, että käyttäjä valitsee haun kannalta sopivimmat termit ja paras laajennuskaava valitsee samat termit. Tulokseksi saatiin, että wpq ja r-lohi kaavojen termeistä yli 60 % esiintyi myös käyttäjän valitsemissa termeissä (Efthimiadis 1995.)

Efthimiadiksen (1995) tutkimusta voidaan kritisoida siitä, että käyttäjän valitsemat laajennustermit eivät välttämättä ole parempia kuin automaattisesti valitut laajennustermit, eikä käyttäjän tekemää termivalintaa pitäisi näin ollen käyttää kriteerinä laajennuskaavojen tehokkuudelle. Tähän tulokseen päätyivät mm. Magennis ja van Rijsbergen (1997) vertaillen dokumenttien valintaan perustuvaa relevanssipalautetta termirelevanssipalautteeseen. Heidän mukaansa käyttäjän valitsemilla laajennustermeillä on mahdollista parantaa haun tuloksellisuutta, mutta tämä vaatii käyttäjien kouluttamista termien valintaan. Relevanssipalautteen tuottamaan hakutehokkuuden kasvuun termien valintaan perustuvilla menetelmillä voi olla hankala päästä. (ks. myös Ruthven ja Lalmas 2003, 35.). Kiinnostavaa Efthimiadiksen (1995) tutkimuksessa on kuitenkin se, että eri kaavat tuottivat suunnilleen samat laajennustermit. Kaavat erosivat toisistaan useimmiten vain järjestyksen osalta, joka sekin oli kuitenkin varsin samankaltainen eri kaavoilla.

### **4.3 Avainten valinta ja painottaminen**

Perinteisesti osittaistämättävissä tiedonhakujärjestelmissä hyödynnetään termifrekvenssiin ( $tf =$  termin esiintymiskerrat dokumentissa) ja käänteiseen dokumenttifrekvenssiin (kaava 11) perustuvaa  $tf*idf$ -kaavaa. Ajatuksena tässä on, että mitä useammin kyselytermi esiintyy dokumentissa, mutta mitä harvemmassa dokumentissa kyselytermi esiintyy, sitä todennäköisemmin dokumentti on relevantti kyselyn kannalta. (Ks. esim. Pirkola, Leppänen ja Järvelin 2002.) Termifrekvenssi voidaan lisäksi normalisoida dokumentin pituudella, jotta pitkät dokumentit eivät saisi liian korkeita arvoja (Salton ja Buckley 1988).

Kaava 11 
$$idf_t = \log \frac{N}{n_t}$$

$N$  = dokumenttien lukumäärä tietokannassa

$n_t$  = dokumenttien lukumäärä, jossa termi  $t$  esiintyy

$tf*idf$  -kaavan avulla dokumenteissa esiintyvillä termeillä voidaan laskea painoja, joiden perusteella dokumentit voidaan listata relevanssin mukaiseen järjestykseen.  $tf*idf$  -kaava ei kuitenkaan tarkoita, että korkean arvon saanut termi olisi hyvä hakuavain, sillä relevantissa dokumentissa hyvä hakuavain voi esiintyä vain kerran, vaikka sillä yleisesti olisikin korkea termifrekvenssi dokumenteissa. (Ks. esim. Pirkola, Leppänen ja Järvelin 2002.)

Tutkimuksessani kiinnostuksen kohteena on Pirkolan, Leppäsen ja Järvelinin (2002) kehittämä RATF -kaava, jonka avulla voidaan valita kyselynlaajennuksessa käytettäviä termejä dokumenteista. Tätä ennen on kuitenkin syytä esitellä Kwokin (1996) kehittämä kyselytermien painotuskaava, johon myös RATF osin perustuu.

#### 4.3.1 Kwokin (1996) kaava

Kwok (1996) testasi noin kuuden sanan mittaisilla kyselyillä kyselytermien painottamisen vaikutusta hakutehokkuuteen. Aluksi hän painotti kyselyitä manuaalisesti toistamalla keskeisimpiä sanoja kyselyssä ja huomasi tämän parantavan hakutehokkuutta. Ajatuksena hänellä oli, että tiedonhakuprosesseissa on tavallista kohdella kyselyitä ja dokumentteja samalla tavalla. Näin on, jos esimerkiksi täsmäytys perustuu dokumenttien ja kyselyiden samankaltaisuuden laskentaan. Ongelmana edellä kuvatun kaltaisessa täsmäytyksessä on se, että lyhyissä kyselyissä ei voida hyödyntää kyselytermien esiintymistiheyden perustuvia painotuksia, sillä lyhyissä kyselyissä termit esiintyvät tavallisesti vain kerran. Tämän pohjalta hän kehitti kaavan, jonka avulla kyselyn termejä voidaan painottaa automaattisesti ilman käyttäjää. Kwokin (1996) kaavan keskeinen huomio on se, että hyvillä hakutermeillä on korkea keskimääräinen termifrekvenssi, eli sana esiintyy dokumentissa monta kertaa, mutta pieni dokumenttifrekvenssi, eli sana ei esiinny kovin monessa dokumentissa.

(Kaava 12) 
$$avtf_k = ColFre_k / DocFre_k$$

Kaavassa 12  $ColFre_k$  on kokoelmataajafrekvenssi, joka laskee termin  $k$  esiintymiskerrat kokoelmassa siten, että myös termin  $k$  toistuva käyttö dokumentissa huomioidaan.  $DocFre_k$  on termin  $k$  dokumenttifrekvenssi, joka laskee, kuinka monessa dokumentissa termi esiintyy, mutta huomio vain termin esiintymisen. Näin laskettu keskiarvo kertoo, että tietyssä kokoelmassa termi esiintyy  $avtf_k$  kertaa dokumentissa keskimäärin. Kaava laskee vähän toisteisille harvinaisille termeille arvoiksi noin 1, vaikka ne voivatkin olla hyödyllisiä sisällönkuvaajia. Samoin yleiset termit tapaavat esiintyä usein niin dokumenteissa kuin kokoelmassa yleisestikin ja myös ne saavat pienehkön  $avtf$ -arvon. Tarkoituksena on antaa korkea paino sellaisille termeille, joilla on suuri  $avtf$ -arvo ja jotka esiintyvät tarpeeksi monessa dokumentissa. Seuraavalla kaavalla  $avyf_k$ -kaavaa muutetaan paremmin tarkoitusta vastaavaksi.

(Kaava 13) 
$$w_k = (avtf_k)^\alpha / \log \max(cutoff, DocFre_k) / norm$$

$norm = \text{normalisointiarvo} \sum_k w_k = 1$

$\alpha$  saa arvon väliltä 1 ja 2. Sen avulla säädellään  $avtf$ :n vaikutusta.

$cutoff$  = kokoelmakohtainen raja-arvo termifrekvenssille

Käänteisen dokumenttifrekvenssin avulla korkeafrekvenssisia termejä voidaan painottaa alas ja matalafrekvenssisia termejä ylös.

Kaavan avulla kyselytermeille lasketaan painot ja kyselyä ”laajennetaan” ilman, että uusia termejä lisätään kyselyyn. Kwok (1996) ei tutkimuksessaan laskenut kaavalla kaikille kyselyn termeille painoja, sillä väärin painotettu termi voi pudottaa tulosta ratkaisevasti. Käytännössä painotettavat termit valitaan niiden termifrekvenssin perusteella. Lisäksi alle neljän sanan pituisiin kyselyihin painotusta ei tehty lainkaan, sillä riski heikentää hakutulosta on tällöin liian suuri. Tutkimuksen mukaan lyhyiden kyselyiden (6-8 sanaa) tuloksellisuutta pystyttiin parantamaan painotuksen avulla. (Kwok 1996.)

#### 4.3.2 RATF

Kwokin (1996) kaavasta johdettua RATF -kaavaa voidaan käyttää termien poimimiseen dokumenteista. Näitä termejä voidaan lisätä haluttu määrä käyttäjän tekemään kyselyyn ja laajentaa kyselyä. RATF -kaavan kehittivät Pirkola, Leppänen ja Järvelin (2002) Kwokin (1995) esittämän



kaavan pohjalta. RATF -kaava perustuu Pirkolan ja Järvelinin aikaisempaan tutkimukseen, jonka mukaan tyypillisesti yhdellä tai kahdella kyselyn avaimista on huomattavasti korkeampi  $cf/df$  ja huomattavasti pienempi dokumenttifrekvenssi ( $df$ ) kuin muilla kyselyn avaimilla.

Kaava laskee sanoille lukuarvoja siten, että mitä harvemmin termi esiintyy tietokannassa ja mitä useammin dokumentissa keskimäärin, sitä suuremman lukuarvon termi saa. RATF perustuu keskiarvojen laskemiseen, eli varsinaista dokumenttikohtaista termifrekvenssiä kaava ei hyödynnä. Suurimman RATF -arvon saaneiden termien oletetaan kuvaavan paremmin dokumentin sisältöä kuin tietokannassa yleisesti esiintyvien termien, jotka ovat saaneet pienemmän arvon. RATF -kaava on seuraavanlainen:

(Kaava 14) 
$$RATF_k = (cf_k / df_k) * 10^3 / \ln(df_k + SP)^p$$

$RATF_k$  = relative average term frequency of the key  $k$ . Eli hakuavaimen  $k$  suhteellinen keskimääräinen termifrekvenssi.

$cf_k$  = hakuavaimen  $k$  esiintymiskerrat dokumenttikokoelmassa (kokoelmafrekvenssi), myös useammat esiintymiset samassa dokumentissa lasketaan

$df_k$  = dokumenttien lukumäärä, jossa hakuavain  $k$  esiintyy

$cf_k / df_k$  = on sama kuin Kwokin (1996) kaavan  $avtf_k$  keskimääräinen termifrekvenssi, joka kertoo, kuinka monta kertaa termi  $k$  esiintyy kokoelman dokumenteissa keskimäärin

$SP$  = kokoelmakohtainen parametrin arvo. Sen avulla alaspainotetaan hyvin harvinaisia sanoja, jotta ne eivät saa liian suurta RATF -arvoa.

$p$  = on kokoelmakohtainen potenssimuuttuja, joka vahvistaa  $SP$ :n vaikutusta.

Kaava antaa siis korkean painon (lukuarvon) sellaisille sanoille, joiden keskimääräinen termifrekvenssi on korkea (eli ne esiintyvät monta kertaa dokumenteissa keskimäärin) ja joilla on alhainen dokumenttifrekvenssi, eli sana ei saisi esiintyä kovin monessa dokumentissa (kuitenkin niin, että harvinaiset sanat eivät saa saada liian suurta painoa). Sanojen saamat RATF-arvot eivät ole dokumenttikohtaisia. Sen sijaan RATF -arvo on riippuvainen dokumenttikokoelmasta ja myös parametrien  $SP$  ja  $p$  arvot on säädettävä kokoelmakohtaisesti.

Keskustalo, Järvelin ja Pirkola (2006) käyttivät menestyksekkäästi RATF -kaavaa relevanssipalautteella toteutetussa kyselynlaajentamisessa. He eivät painottaneet kyselytermejä Kwokin (1996) tapaan, vaan käyttivät kaavaa laajennustermien valintaan. Relevanssipalautetta varten valittiin ensin sopiva määrä dokumentteja, jonka jälkeen jokaiselle dokumenteissa esiintyneelle termille laskettiin RATF -arvo. Sitten relevanteiksi valituista dokumenteista valittiin 50 suurimman RATF -arvon saanutta termiä. Tämän jälkeen dokumenttien RATF -listat yhdistettiin ja laajennukseen valittiin 30 termiä sen mukaan, kuinka monessa dokumentissa ne esiintyivät ja kuinka korkean RATF -arvon termit saivat. Termejä ei siis valittu laajennukseen pelkän RATF -arvon perusteella. Oletuksena oli, että termi joka esiintyy useassa (oletetussa) relevantissa dokumentissa riittävän suurella RATF - arvolla, on todennäköisemmin hyödyllinen kuin sellainen termi, joka esiintyy suurella RATF -arvolla vain yhdessä dokumentissa. Tutkimukseen palataan vielä palautedokumenttien määrää käsittelevässä luvussa 4.5.

RATF -kaavan käyttöä relevanssipalautteen laajennustermien valinnassa voidaan perustella myös teorian pohjalta. Monissa muissa automaattista kyselynlaajentamista ja laajennustermien valintaa käsittelevissä tutkimuksissa laajennustermien valinta perustuu termien esiintymistiheyteen hakutermin kanssa. Hakutermin ja laajennustermien korrelaation perustuvan laajennuksen ongelmana on kuitenkin se, että hakutermit ovat usein melko yleisiä termejä, jolloin myös laajennustermeiksi valikoituu melko yleisiä termejä. Yleiset termit erottelevat relevanteja ja epärelevanteja dokumentteja huonosti, minkä seurauksena laajennus epäonnistuu. (Ks. Efthimiadis 1996, 151.)

#### ***4.4 Palautteeseen valittavien laajennustermien määrä***

Kun relevanssipalautetta varten on valittu menetelmä termien poimimiseksi dokumenteista, on seuraavaksi säädettävä menetelmä toimimaan optimaalisella tavalla. Tähän liittyy kysymys siitä, kuinka monta termiä laajennettuun kyselyyn pitäisi valita. Seuraavaksi esittelen joitakin tutkimuksia, joissa tähän on perehdytty.

Buckley, Salton ja Allan (1994) tutkivat relevanssipalautteessa käytettyjen termien määrän vaikutusta hakutehokkuuteen vektorimalliin perustuvassa hakujärjestelmässä. Hei poimivat dokumenteista 0-4000 termiä laajennukseen. Termit valittiin sen mukaan, kuinka monessa relevantissa dokumentissa termi esiintyi. Relevanssiarvioitavien dokumenttien määrää vaihdeltiin 5

ja 5000 välillä. Laajennettua kyselyä painotettiin muunnellulla Rocchion kaavalla, joka vastaa pääpiirteissään luvussa 4.1 esitettyä kaavaa 4. Erotuksena muunnellussa kaavassa on se, että kaikki tulosjoukon ulkopuolelle jäävät dokumentit oletetaan epärelevanteiksi. Ajatuksena tässä on, että tulosjoukon epärelevantit dokumentit eivät edusta kattavasti epärelevanttien dokumenttien joukkoa. Kaava laskee matalan painokertoimen termeille, jotka esiintyvät harvoin dokumenteissa. (Buckley, Salton ja Allan 1994.)

Tutkimuksen mukaan parhaiten hakutehokkuuteen vaikutti kyselyn laajentaminen 500 termillä. Termien kasvattaminen tätä suuremmaksi ei parantanut enää tuloksellisuutta. Kun termejä valittiin viidestä relevantista dokumentista, hakutehokkuutta voitiin kasvattaa 19 prosenttiyksikköä laajentamattomaan kyselyyn verrattuna. Tutkimuksessa onnistuttiin kasvattamaan hakutehokkuutta jopa 38 prosenttiyksikköä, mutta tällöin palautetta varten piti käydä läpi 2500 dokumenttia. Palautedokumenttien määrää nostamalla pystyttiin näin ollen kasvattamaan laajennetun kyselyn tuloksellisuutta varsin selvästi. Luvussa 5 esitetyn relevanssipalautteeseen liittyvän ongelman dokumenttien ”kaksoisesiintymisistä” alkuperäisessä ja laajennetuissa haussa he välttivät siten, että he käyttivät kahta eri dokumenttikokoelmaa, joissa ei ollut yhteisiä dokumentteja. Tämä oli mahdollista, sillä heillä oli käytössään melko suuri, satoja tuhansia dokumentteja sisältävä TREC -dokumenttikanta. (Buckley, Salton ja Allan 1994.)

Edellisessä luvussa esitellyssä Keskustalon, Järvelinin ja Pirkolan (2006) tutkimuksessa päästiin hyviin tuloksiin vain 30 laajennustermillä. Näiden tutkimusten yhteenvetona voidaan sanoa, että mitään yleispätevää laajennustermien lukumäärästä ei voida esittää. Haun tuloksellisuus on parantunut niin 30:llä kuin 500:llä laajennustermillä. Termien lukumäärä riippuu käytössä olevasta hakujärjestelmästä ja menetelmästä, jolla laajennustermit valitaan. Laajennukseen valittavien dokumenttien lukumäärää puolestaan ei voida kovin suureksi kasvattaa. Tätä rajoittaa hakuaiheeseen olevien relevanttien dokumenttien lukumäärän lisäksi käyttäjän kiinnostuneisuus antaa palautetta. Lisäksi laajennustermien määrään vaikuttaa käytettävissä oleva dokumenttikokoelma ja kyselyt (ks. esim. Ruthven ja Lalmas 2003, 31).

## **4.5 Palautteeseen valittavien dokumenttien määrä**

Toinen relevanssipalautteen ja erityisesti pseudorelevanssipalautteen toimivuuteen vaikuttava tekijä on laajennustermien määrän lisäksi se, montako dokumenttia palautteeseen pitäisi ottaa mukaan. Edellisessä luvussa esitellyn Buckleyn, Saltonin ja Allanin (1994) tutkimuksen perusteella on oletettavaa, että useasta relevantista dokumentista annettu palaute tuottaa parempia laajennustermejä ja näin ollen myös tehokkaamman laajennuksen kuin esimerkiksi yhden relevantin dokumentin palaute.

Laboratorio-olosuhteissa palautedokumenttien määrää voidaan kasvattaa mielivaltaisesti ja päästä erittäin hyviin hakutuloksiin. Harva käyttäjä kuitenkaan viitsii seuloa satoja tai jopa tuhansia dokumentteja, vaikka voisikin näin löytää taas uusia relevantteja dokumentteja. Käyttäjät kuitenkin vaihtelevat preferensseiltään ja tiedontarpeiltaan. Siinä missä yksi ei halua antaa relevanssipalautetta lainkaan voi toinen olla halukas selaamaan läpi pidemmänkin tulosjoukon. Relevanssipalautteen hyödyllisyyttä voidaan tutkia laboratorio-olosuhteissa mallintamalla erilaisia käyttäjätilanteita.

Lähtökohtana monissa relevanssipalautetta käsittelevissä tutkimuksissa on että, aluksi otetaan tulosjoukon kärjestä 10 - 20 dokumenttia. Näistä dokumenteista valitaan relevantteja dokumentteja ja valittujen relevanttien dokumenttien perusteella tehdään uusi kysely. (Ks. esim. Efthimiadis 1996, 136.) Laboratoriotutkimuksissa, joihin ei osallistu oikeita käyttäjiä, tämä voidaan toteuttaa seuraavalla tavalla: oletetaan, että käyttäjä valitsee tulosjoukosta tietyn määrän relevantteja dokumentteja, joiden perusteella laajennus tehdään. Näin käyttäjätilanteita voidaan simuloida esimerkiksi sen mukaan, kuinka halukkaita käyttäjät ovat antamaan palautetta tai millä relevanssikriteereillä he valitsevat relevantteja dokumentteja palautteeseen.

Edellä esitetyn kaltaista käyttäjätilanteiden simulointia käyttivät mm. Keskustalo, Järvelin ja Pirkola (2006). He tutkivat RATF:lla toteutetun relevanssipalautteen vaikutusta haun tuloksellisuuteen. Heillä oli käytössään neliportaisesti relevanssiarvioitu dokumenttikokoelma, joissa dokumentit oli jaettu epärelevantteihin, marginaalisesti relevantteihin, osittain relevantteihin ja erittäin relevantteihin. He selvittivät kuinka relevanssipalautteen määrä ja laatu vaikuttaa hakutehokkuuteen. Lisäksi he vertailivat pseudorelevanssipalautetta ja eri käyttäjätilanteissa toteutettua relevanssipalautetta keskenään. He simuloivat käyttäjätilanteita ensinnäkin sen mukaan, minkä asteista relevanssia käyttäjät odottivat dokumenteilta. Liberaaleilla relevanssikriteereillä

toimivalle käyttäjälle kelpasivat kaikki vähänkin relevantit dokumentit. Keskitiukoilla kriteereillä toimivalle käyttäjälle puolestaan kelpasivat muut paitsi marginaalisesti relevantit dokumentit ja tiukoilla relevanssikriteereillä toimiva käyttäjä kelpuutti mukaan vain kaikkein relevanteimmat dokumentit. Käyttäjätilanteiden simuloinnissa hyödynnettiin lisäksi tuloslistan pituutta siten, että laiska käyttäjä suostui antamaan palautetta vain ensimmäisen kymmenen dokumentin joukosta, kun taas viitseliäämpi käyttäjä selasi 30 ensimmäistä dokumenttia. Kolmanneksi käyttäjätilanteita mallinnettiin sen mukaan, kuinka monesta dokumentista käyttäjät halusivat antaa relevanssipalautetta. Hakulausekkeet tutkimuksessa muodostettiin automaattisesti TRECUTA hakuaiheiden kokoelmasta (ks. dokumenttikokoelmasta luvusta 8.3) otsikkokentän ja kuvailukentän termeistä. (Keskustalo, Järvelin ja Pirkola 2006.)

Tutkimuksen mukaan relevanssipalautte parantaa haun tuloksellisuutta. Paras tuloksellisuus saavutetaan, kun palautetta annetaan vähintään viidestä erittäin relevantista dokumentista, jolloin haun tuloksellisuus on n. 37 prosenttiyksikköä parempi kuin alkuperäisessä kyselyssä. Tämän saavuttaakseen on käyttäjän käytävä läpi n. 30 tulosjoukon kärkidokumenttia. Lisäksi tutkimuksen mukaan pieni mutta korkeatasoinen (hakija kelpuuttaa vain erittäin relevantit dokumentit hyödyllisiksi) relevanssipalautte toimii laajaa, mutta heikkotasoisista (hakija kelpuuttaa myös heikommat dokumentit relevanteiksi) relevanssipalautetta paremmin. (Keskustalo, Järvelin ja Pirkola 2006.) Relevanssipalautteen antamista yli 30 dokumentista ei tutkimuksessa testattu, mutta tulosjoukon kasvattaminen 30 dokumenttiin vaikutti kasvattavan hakutehokkuutta.

Pseudorelevanssipalautteessa tilanne ei kuitenkaan ole yhtä selvä. Palautedokumenttien määrää kasvattamalla ei välttämättä ole positiivista vaikutusta hakutehokkuuteen. Kun palautedokumentit valitaan sokkona tulosjoukon kärjestä, mukana pitäisi olla riittävästi relevantteja dokumentteja, mutta kuinka suuressa tulosjoukossa on tarpeeksi relevantteja dokumentteja keskimäärin, jotta pseudorelevanssipalautte toimisi tehokkaasti? Tässä luvussa esitellään pari tutkimusta, joissa asiaa on käsitelty.

Lundquist, Grossman ja Frieder (1997) vertailivat erilaisia pseudorelevanssipalautetekniikoita ja pyrkivät selvittämään, kuinka yksittäinen relevanssipalautetekniikka vaikuttaa relevanssipalautteen toimivuuteen. Heillä oli käytössään vektorimalliin perustuva tiedonhakujärjestelmä ja 50 TREC-4 hakuaihetta. He selvittivät ensinnäkin, kuinka monta dokumenttia tuloslistan kärjestä relevanssipalautteeseen tulee ottaa mukaan parhaan hyödyn saamiseksi. Parhaat saanti- ja tarkkuusluvut he saivat, kun he käyttivät laajennukseen 5-20 kärkidokumenttia. Dokumenttien

lisääminen tai vähentäminen huononsi hakutuloksen tarkkuutta. Toisekseen he testasivat, kuinka monta hakutermiä laajennukseen kannattaa poimia dokumenteista. Tätä tarkoitusta varten he käyttivät 50 lyhennettyä TREC-5 kyselyä. He päätyivät parhaaseen tulokseen, kun kyselyitä laajennettiin 10-20 termillä tai fraasilla.

Keskustalo, Järvelin ja Pirkola (2006) testasivat myös pseudorelevanssipalautteen toimivuutta kyselynlaajennukseen (tarkempi selostus tutkimuksesta edellisessä luvussa). He valitsivat laajennustermejä 1, 5, 10 ja 30 kärkidokumentista ja lisäsivät termit alkuperäiseen kyselyyn. Heidän tutkimuksesta ei kuitenkaan voi suoraan päätellä, kuinka monta dokumenttia laajennukseen pitäisi valita. Tutkimuksessa käytetty neliportainen relevanssiarviointi aiheutti sen, että optimaalinen laajennusdokumenttien määrä vaihteli sen mukaan, millä kriteereillä käyttäjä arvioi dokumentit relevantiksi. Tiukoilla relevanssikriteereillä toiminut käyttäjä ei juurikaan hyötynyt pseudorelevanssipalautteesta. Haun tuloksellisuutta onnistuttiin kuitenkin parantamaan keskimäärin 1,8 prosenttiyksikköä, kun laajennus tehtiin yhden dokumentin perusteella. Liberaaleilla relevanssikriteereillä toiminut käyttäjä hyötyi 3,8 prosenttiyksikköä laajennuksesta kun palaute muodostettiin 10 kärkidokumentin perusteella. Tutkimuksesta on kuitenkin vaikea päätellä, kuinka monesta dokumentista pseudorelevanssipalautetta kannattaisi antaa, sillä erot hakutehokkuuden suhteen ovat varsin pieniä, eikä tulosten tilastollista merkitsevyyttä testattu.

Sekä Buckley, Saltonin ja Allanin (1994) että Keskustalon, Järvelinin ja Pirkolan (2006) tutkimukset viittaavat siihen, että relevanssipalautteen avulla löydetään sitä enemmän relevantteja dokumentteja, mitä suuremmasta tulosjoukosta palautetta annetaan. Tehokas relevanssipalautejärjestelmä toimii kuitenkin menestyksekkäästi jo pienillä palautedokumenttimäärillä, sillä käyttäjän ei voida olettaa seulovan satojen dokumenttien pituisia tulosjoukkoja. Pseudorelevanssipalautteessa sen sijaan tulosjoukon kokoa voitaisiin kasvattaa periaatteessa mielivaltaisesti, sillä tästä ei koidu käyttäjälle lisävaivaa. Tässä esiteltyt Keskustalon, Järvelinin ja Pirkolan (2006) ja Lundquistin, Grossmanin ja Friederin (1997) tutkimukset kuitenkin viittaavat siihen, että tulosjoukon kasvattaminen kovin suureksi ei paranna pseudorelevanssipalautteen tehokkuutta, vaikkakaan kovin massiivisia laajennuksia ei edes testattu. Jo 5-10 dokumentista tehdyllä palautteella näyttäisi olevan mahdollista parantaa hakutehokkuutta.

## 5. Palautekyselyn muodostaminen

Kun laajennustermit on valittu relevanssipalautteen avulla, ne on liitettävä vielä alkuperäiseen kyselyyn. Kyselytermien liittäminen toisiinsa tapahtuu rakenteen avulla. Tässä luvussa selvitän aluksi, mitä kyselyn rakenteella tarkoitetaan, jonka jälkeen esittelen lyhyesti tutkimuksen, jossa rakenteista kyselyä on sovellettu relevanssipalautteeseen.

Kyselyn rakenteella tarkoitetaan hakutermien välistä suhdetta, joka voidaan ilmaista kyselykielessä käytettävillä operaattoreilla tai hakuavainten painotuksilla. Boolean logiikkaa noudattavissa hakumallissa termien välisiä suhteita voidaan ilmaista AND ja OR -operaattoreilla, kun taas osittaistämättävissä malleissa hakuavainten painottaminen on tavallisempaa. Hakuavainten painottaminen on syytä erottaa dokumenttiteksteissä esiintyvien sanojen painotuksesta. Osittaistämättävissä hakujärjestelmissä dokumenttien relevanssilajittelu perustuu dokumenttien teksteissä esiintyville sanoille laskettuihin painoihin, kun taas hakuavainten painotuksessa on kysymys siitä, että joitakin hakulauseen termejä halutaan painottaa enemmän kuin toisia. Avainten panot voivat olla joko binäärisiä, jolloin paino voi saada arvon yksi tai nolla riippuen siitä, esiintyykö termi dokumentissa tai kyselyssä vai ei. Ei-binäärisessä termien painotuksessa painojen laskenta perustuu usein luvussa 4.3 esitettyyn  $tf*idf$  -kaavaan tai siitä johdettuun kaavaan. Vaihtoehtoisesti käyttäjä voi itse antaa kyselyn termeille haluamiaan painoja.

Rakenteisella kyselyllä viitataan usein Boolean logiikan avulla muodostettuun kyselyyn. Relevanssipalautetta puolestaan sovelletaan tavallisesti osittaistämättäviin hakujärjestelmiin, joissa rakenteisia kyselyitä sovelletaan harvemmin ainakaan niin, että loppukäyttäjä voisi kyselyn rakenteeseen vaikuttaa. Ehkä tästä syystä relevanssipalautetta käsittelevissä tutkimuksissa ei juurikaan käsitellä kyselyn rakennetta.

Keskustalo, Järvelin ja Pirkola (2006) tekevät kuitenkin tässä suhteessa poikkeuksen. He tutkivat relevanssipalautetta osittaistämättävällä InQuery -tiedonhakuajärjestelmällä, jossa on mahdollista muodostaa rakenteisia kyselyitä. He muodostivat alkuperäisen kyselyn automaattisesti otsikkokentän sanoilla siten, että ensin he poistivat stop -sanalistan avulla kyselyn kannalta hyödyttömät termit ja tämän jälkeen he perusmuotoistivat sanat. Perusmuotoistetut sanat liitettiin toisiinsa InQueryn  $\#sum$  -operaattorilla, joka laskee jokaiselle sanalle yhtä suuren painon. Mikäli perusmuotoistettu sana oli monitulkintainen, kaikki perusmuotoistamisen seurauksena saadut muodot liitettiin toisiinsa  $\#syn$  operaattorilla, joka tulkitsee kaikki operaattorilla yhdistetyt sanat

yhden sanan esiintymäksi. Tällöin alkuperäinen kysely saa rakenteen  $\#sum(\#syn(key1, key2, \dots), \#syn(\dots, keyn, \dots))$ . Kyselynlaajennukseen käytetyt sanat muodostivat rakenteen  $\#sum(keye1, keye2, \dots)$  ja uusi laajennettu kysely noudatti rakennetta  $\#sum(\#sum(\#syn(key1, key2, \dots), \#syn(\dots, keyn), \dots) \#sum(keye1, keye2, \dots))$ .

Toinen vaihtoehto edellä esitetyn kaltaiselle rakenteelle voisi olla hakufasettien painottaminen esimerkiksi siten, että käyttäjän tekemälle alkuperäiselle kyselylle annetaan suurempi paino kuin laajennustermien muodostamalle fasetille. Relevanssipalautekyselyyn ei kuitenkaan ole mahdollista tuottaa kovin monimutkaista rakennetta, mikäli prosessi halutaan pitää käyttäjälle mahdollisimman yksinkertaisena. Rakenteisen kyselyn tuottaminen edellyttää termien välisten suhteiden tunnistamista, mikä ei onnistu kovin helposti ilman käyttäjää.



## 6. Relevanssipalautteen tehokkuuden mittaaminen

Kun relevanssipalautteen avulla tehty laajennettu kysely on muodostettu ja tarkoitus olisi päästä testaamaan relevanssipalautteen toimintaa, on ratkaistavana vielä yksi ongelma. Miten relevanssipalautteen tuloksellisuutta mitataan?

Relevanssipalautteen tarkoituksena on laajentaa kyselyä käyttäjän valitsemien relevanttien dokumenttien perusteella ja parantaa näin haun tuloksellisuutta. Tästä seuraa helposti se, että kun relevanteiksi arvioiduista dokumenteista valitaan laajennustermejä, samat relevantit dokumentit nousevat tulosjoukon kärkeen myös uudessa tuloslistauksessa. Kun oletetaan, että käyttäjää kiinnostavat vain sellaiset relevantit dokumentit, joita hän ei ole vielä löytänyt, ei uusi tuloslista olekaan vertailukelpoinen alkuperäisen tulosjoukon kanssa. Näin ollen varsinaisen relevanssipalautteen aikaansaama hakutehokkuuden muutosta ei voidakaan todentaa. Chang, Cirillo ja Razon (1971) esittävät kolme tapaa, jolla edellä kuvatun kaltainen ongelma voidaan välttää. (Ks. myös Ruthven ja Lalmas 2003, 17.)

Yksi tapa parantaa relevanssipalautteen mittauksen luotettavuutta on käyttää ns. jäännöskokoelmaa (residual collection). Tässä menetelmässä käyttäjän relevanssipalautteeseen valitsemat dokumentit poistetaan dokumenttikokoelmasta. Tämän jälkeen sekä alkuperäinen laajentamaton haku, että relevanssipalautteen avulla laajennettu kysely tehdään jäännöskokoelmaan. Näin tuloksista saadaan vertailukelpoisia. Mikäli alkuperäistä kyselyä ei tehdä uudelleen jäännöskokoelmaan, ei laajennetun kyselyn tuloksellisuus ole täysin vertailukelpoinen alkuperäisen haun kanssa, sillä relevantteja dokumentteja on jäännöskokoelmassa vähemmän, jolloin mahdollisuudet parempaan hakutulokseen ovat myös heikommat. Baeza-Yates ja Ribeiro-Neto (1999) ehdottavat ratkaisuksi tähän, että vertailua alkuperäisen ja laajennetun haun välillä ei pitäisi tehdä lainkaan, vaan tarkoituksenmukaisinta olisi verrata eri keskenään eri relevanssipalautestrategioita.

Toinen tapa mitata relevanssipalautteen vaikutusta hakutehokkuuteen on tulosjoukon jäädyttäminen (freezing). Tässä menetelmässä käyttäjän relevanssipalautteeseen valitsemat dokumentit sekä valitsematta jättämät epärelevantit dokumentit (esimerkiksi ensimmäisen kymmenen dokumentin joukosta) jäädytetään paikoilleen. Tämän jälkeen jäljelle jääneet dokumentit järjestetään uudelleen laajennetulla kyselyllä. Tämänkaltaisia relevanssipalauttekerroksia voidaan tehdä useita, jolloin jäädytetyn tuloslistan pituus kasvaa, kun

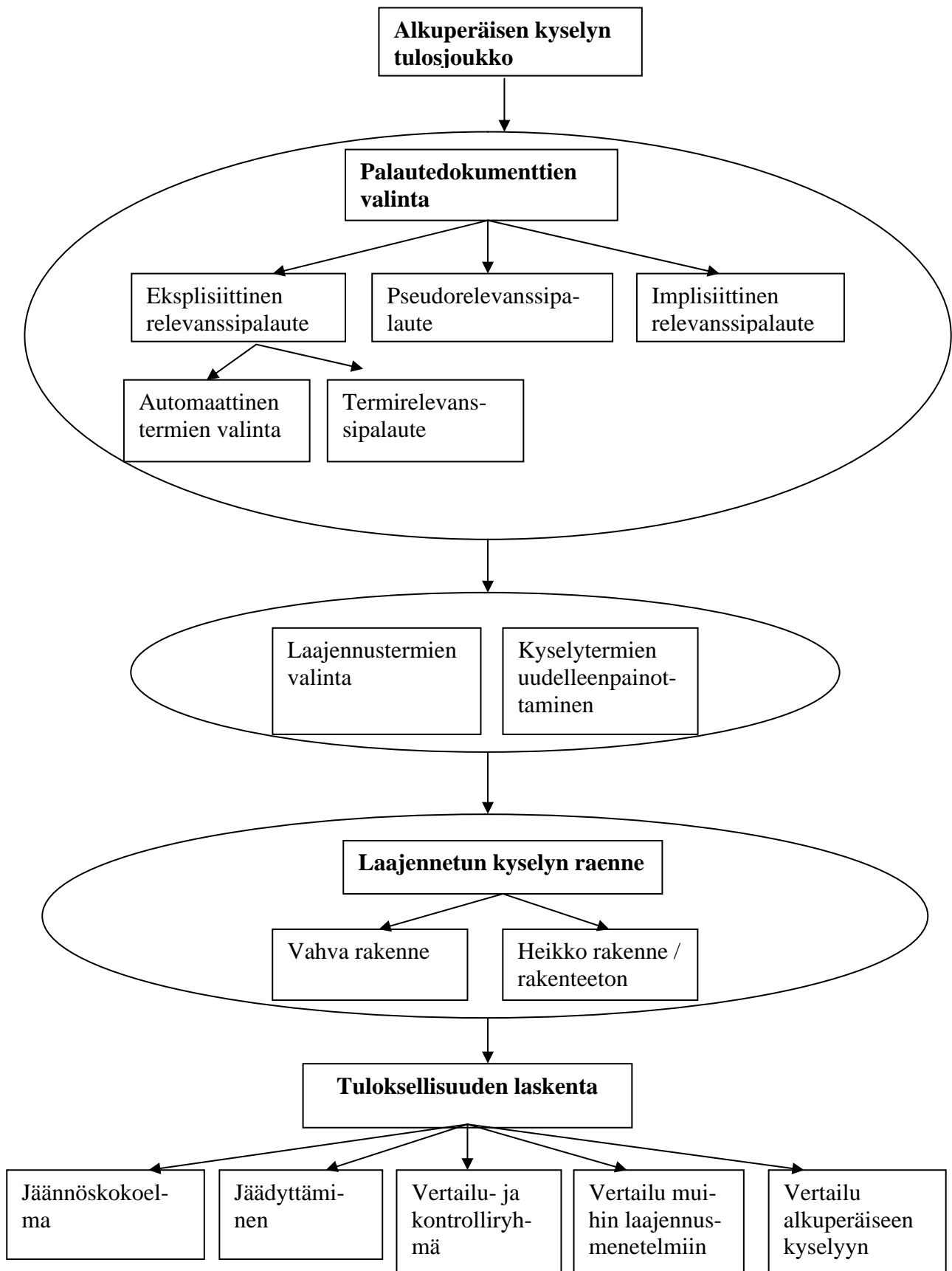
uusi tulosjoukko liitetään aina aikaisempaan jäädytettyyn tuloslistaan. Laajennetun haun tuloksellisuus voidaan näin ollen laskea koko jäädytetystä tulosjoukosta kaikkien relevanssipalauttekerrosten jälkeen. Menetelmää voidaan muunnella myös siten, että jäädytetään ainoastaan relevantit dokumentit. Jäädyttämisen avulla muodostetaan tilanne, jossa käyttäjä voi joko antaa relevanssipalautetta tai jatkaa tuloslistan selailua eteenpäin uusien relevanttien dokumenttien toivossa. Tarkoituksena on mitata, kumpi menetelmä tuottaa paremman hakutehokkuuden.

Kolmas keino relevanssipalautteen tuloksellisuuden mittaamiseksi on käyttää testi- ja kontrolliryhmää. Tässä menetelmässä dokumenttikanta jaetaan kahteen osaan siten, että relevantit dokumentit jakautuvat tasaisesti molempien dokumenttikantojen kesken. Käyttäjän tekemä alkuperäinen kysely suoritetaan testikokoelmaan ja palautteen avulla laajennettu kysely suoritetaan kontrollikokoelmaan. Näin relevanssipalautteen tehokkuutta voidaan verrata suoraan alkuperäisen kyselyn tehokkuuteen. Ongelmalliseksi tässä menetelmässä muodostuu ensinnäkin dokumenttikannan jakaminen sattumanvaraisesti tasan siten, että relevantit dokumentit jakautuvat tasaisesti molempiin kokoelmiin. (Ks. esim. Ruthven ja Lalmas 2003, 18 ja Chang, Cirillo ja Razon 1971.) Toisen ongelman saattaa aiheuttaa se, että dokumenttikokoelman täytyy olla tarpeeksi suuri, jotta se voidaan jakaa kahtia.

## **7. Yhteenvetoa teoriaosuudesta**

Relevanssipalaute on melko suoraviivainen menetelmä, jolla voidaan laajentaa kyselyä. Kaikilla teoriaosiossa esitellyillä relevanssipalautemenetelmillä haun tuloksellisuutta on onnistuttu parantamaan. Relevanssipalaute on lisäksi helposti toteutettavissa niin vektorimalliin- kuin todennäköisyysmalliinkin perustuvissa hakujärjestelmissä. Tiivistettynä relevanssipalautteen vaiheet ja niiden vaihtoehtoiset toteutustavat voidaan esittää kuvan 4 mukaisesti:

**Kuva 4. Relevanssipalautetutkimuksen toteuttamistavat**



Relevanssipalautteen lähtökohtana on tiedonhakujärjestelmän käyttäjälle palauttama tulosjoukko, jonka dokumenttien perusteella kyselyä laajennetaan. Relevanssipalautte voidaan toteuttaa kuvan 4 mukaisesti ensinnäkin eksplisiittisenä, jolloin käyttäjä valitsee palautedokumentit. Tällöin on edelleen päätettävä, osallistuuko käyttäjä myös laajennustermien valintaan, jolloin kyseessä on termirelevanssipalautte, vai toteuttaako hakujärjestelmä termien valinnan automaattisesti. Automaattinen laajennustermien valinta voidaan toteuttaa esimerkiksi tässä tutkimuksessa käytössä olevalla RATF -kaavalla. Myös termirelevanssipalautteessa voitaisiin hyödyntää RATF:a siten, että kaavalla poimittaisiin relevanteista dokumenteista termejä ja näistä termeistä käyttäjä poimisi haluamansa laajennustermit. Toinen vaihtoehto relevanssipalautteen toteuttamiseksi on valita automaattisesti vakiomäärä dokumentteja tulosjoukon kärjestä, jolloin kyseessä on pseudorelevanssipalautte. Se, montako dokumenttia tulosjoukon kärjestä kannattaa valita optimaaliseen hakutulokseen pääsemiseksi, riippuu osittain relevanssipalauttealgoritmistä ja on yksi tutkimusongelma. Kolmanneksi relevanssipalautte voidaan toteuttaa implisiittisenä, jolloin järjestelmä tarkkailee käyttäjää ja tekee päätelmiä relevanssista. Tämä voidaan toteuttaa esimerkiksi mittaamalla käyttäjän viettämä aika dokumentin parissa.

Termit voidaan valita dokumenteista esimerkiksi luvussa 3.2 esitetyn *wpq* -painotuskaavaan avulla tai luvussa 3.3.2 esitetyn RATF:n avulla. Yksinkertaisimmillaan järjestelmä valitsee termit dokumenteista ja muodostaa uuden kyselyn, jolloin käyttäjän harteille jää ainoastaan dokumenttien valinta. Se montako termiä palautteeseen pitää valita, on yksi relevanssipalauttetutkimuksen osa-alue. Vaihtoehtoisesti termejä ei valita lainkaan, vaan käyttäjän tekemän kyselyn termit painotetaan uudelleen valituissa dokumenteissa esiintyvien termien perusteella. Uusien termien valintaa ja painotusta voidaan käyttää myös samanaikaisesti.

Kun laajennustermit on valittu dokumenteista, on seuraavaksi päätettävä kuinka termit lisätään alkuperäiseen kyselyyn. Tässä vaiheessa voidaan valita painotetaanko alkuperäistä kyselyä enemmän kuin laajennusta ja minkälainen rakenne termeillä on suhteessa toisiinsa. Vahvarakenteisessa kyselyssä termien välisiä suhteita voidaan tunnistaa esimerkiksi perusmuotoistamisen avulla. Lisäksi alkuperäisen kyselyn termejä voidaan painottaa eri tavoin kuin laajennustermejä. Heikkorakenteisessa tai rakenteettomassa kyselyssä kaikkia hakutermejä käsitellään samanarvoisina.

Kun uusi kysely on saatu muodostettua, on jäljellä vielä hakutehokkuuden mittaaminen. Samojen dokumenttien esiintyminen uudelleen laajennetun kyselyn tulosjoukossa voidaan estää tekemällä

laajennettu kysely jäännöskokoelmaan, käyttämällä tuloslistan jäädytystä tai käyttämällä kokonaan uutta dokumenttikokoelmaa. Dokumenttien kaksoisesiintymistä ei tarvitse huomioida, mikäli eri relevanssipalautemenetelmiä verrataan vain keskenään. Relevanssipalautteen tehokkuuden suora vertailu alkuperäiseen kyselyyn puolestaan sopii parhaiten pseudorelevanssipalautteen tapauksessa. Palautekierroksia voidaan tarvittaessa tehdä useita, jolloin laajennetun kyselyn tuottamasta tulosjoukosta voidaan valita uusia relevantteja dokumentteja uutta relevanssipalautetta varten.

## 8. Tutkimuskysymykset ja koeasetelma

Tässä luvussa esittelen aluksi tutkimuskysymykset ja niihin liittyvät hypoteesit. Tämän jälkeen kerron tarkemmin kyselyiden suunnittelusta, testikokoelmasta ja hakujärjestelmästä. Luvussa 8.5 esittelen tulosten analysointiin ja evaluointiin käytetyt menetelmät. Lopuksi esittelen vielä esitutkimuksen ja koeasetelman.

### 8.1 Tutkimuskysymykset

Tutkimuksessa pyritään vastaamaan seuraaviin tutkimuskysymyksiin:

#### 1. Käyttäjän relevanssipalautteen tuloksellisuus

##### 1.1 Miten aloituskyselyn onnistuneisuus vaikuttaa relevanssipalautteen tuloksellisuuteen?

Oletuksena on, että heikkotasoinen kysely hyötyy relevanssipalautteesta ainakin suhteellisesti enemmän kuin parempitasoinen kysely. Jos kysely onnistuu huonosti, myös parantamisen varaa on enemmän. (Ks. esim. Salton ja Buckley 1990.)

1.2. Kuinka monta dokumenttia relevanssipalautteeseen pitäisi valita parhaan hakutuloksen saavuttamiseksi? Hypoteesina on, että useista dokumenteista annettu relevanssipalaute tuottaa myös tuloksekkaan palautekyselyn (Ks. esim. Buckley, Salton ja Allan 1994; Keskustalo, Järvelin ja Pirkola 2006.)

1.3 Minkä tasoisesti relevantteja dokumentteja kannattaa valita relevanssipalautteeseen, jotta palautteesta saadaan maksimaalinen hyöty? Keskustalon, Järvelinin ja Pirkolan (2006) tutkimuksessa erittäin relevanteista dokumenteista annettu relevanssipalaute toimii paremmin kuin palaute, jossa on mukana myös vähemmän relevantteja dokumentteja. Toisaalta, jos kysely onnistuu heikosti, palautetta ei ole aina mahdollista antaa vain relevanteimmista dokumenteista. Tällöin heikosti onnistuneeseen kyselyyn kannattaa ehkä antaa palautetta myös marginaalisesti relevanteista dokumenteista.

## 2. Pseudorelevanssipalautteen tuloksellisuus

2.1 Miten aloituskyselyn onnistuneisuus vaikuttaa pseudorelevanssipalautteen tuloksellisuuteen? Oletuksena on, että tulosjoukossa pitää olla riittävästi relevantteja dokumentteja, jotta pseudorelevanssipalaute onnistuu (ks. esim. Ruthven ja Lalmas 2003). Heikkotasaisen aloituskyselyn voi näin ollen olettaa heikentyvän entisestään pseudorelevanssipalautteella tehdyllä kyselynlaajennuksella, mutta hyvätasoiseen kyselyyn vaikutuksen olettaisi olevan suotuisa.

2.2 Riippuuko pseudorelevanssipalautteen toimivuus palautteeseen valittujen dokumenttien lukumäärästä? Lundquist, Grossman ja Frieder (1997) esittävät, että paras tuloksellisuus saadaan kun kyselyä laajennetaan 5-20 dokumentilla. Tämän pohjalta voidaan ennakoida, että myös tässä tutkimuksessa palaute 5 tai 10 dokumentista tuottaa paremman hakutehokkuuden kuin palaute yhdestä dokumentista.

## **8.2 Testikokoelma**

Tutkimuksessa käytetään sanomalehtiartikkeleita sisältävää TRECUTA -dokumenttikokoelmaa. Kokoelmassa on noin 550 000 dokumenttia ja se sisältää mm. Financial Times- ja Los Angeles Times -lehtien artikkeleita. Kokoelmaan on olemassa 41 hakuaihetta, jotka on poimittu TREC-7 ja TREC-8 aiheista. (ks. lisää Sormunen 2002a.) Kaikkia 41 aiheita ei kuitenkaan voitu ottaa mukaan sen vuoksi, että relevantteja dokumentteja ei ollut riittävästi aiheita kohden. Kuuden aiheen pois jättäminen oli välttämätöntä, sillä tutkimuskohteena ovat eritasoisesti onnistuneet kyselyt, ja mikäli relevantteja dokumentteja on vain vähän, ei kyselyn onnistuneisuuden välille saada tarpeeksi suuria eroja. Niinpä tutkimusta varten käytössä on 35 hakuaihetta ja hakuaiheisiin neliportaisesti relevanssiarvioituja dokumentteja.

Kokoelman dokumentit on arvioitu relevanteiksi asteikolla 0-3, jossa 0 tarkoittaa epärelevanttia dokumenttia ja 1 marginaalisesti relevanttia dokumenttia, jolloin dokumentissa lähinnä viitataan hakuaiheeseen. 2 tarkoittaa kohtuullisen relevanttia dokumenttia, joka käsittelee hakuaihetta, mutta ei kuitenkaan kovin kattavasti. 3 tarkoittaa erittäin relevanttia dokumenttia, joka käsittelee hakuaihetta kattavasti. (Sormunen 2002a.)



### 8.3 Hakujärjestelmä

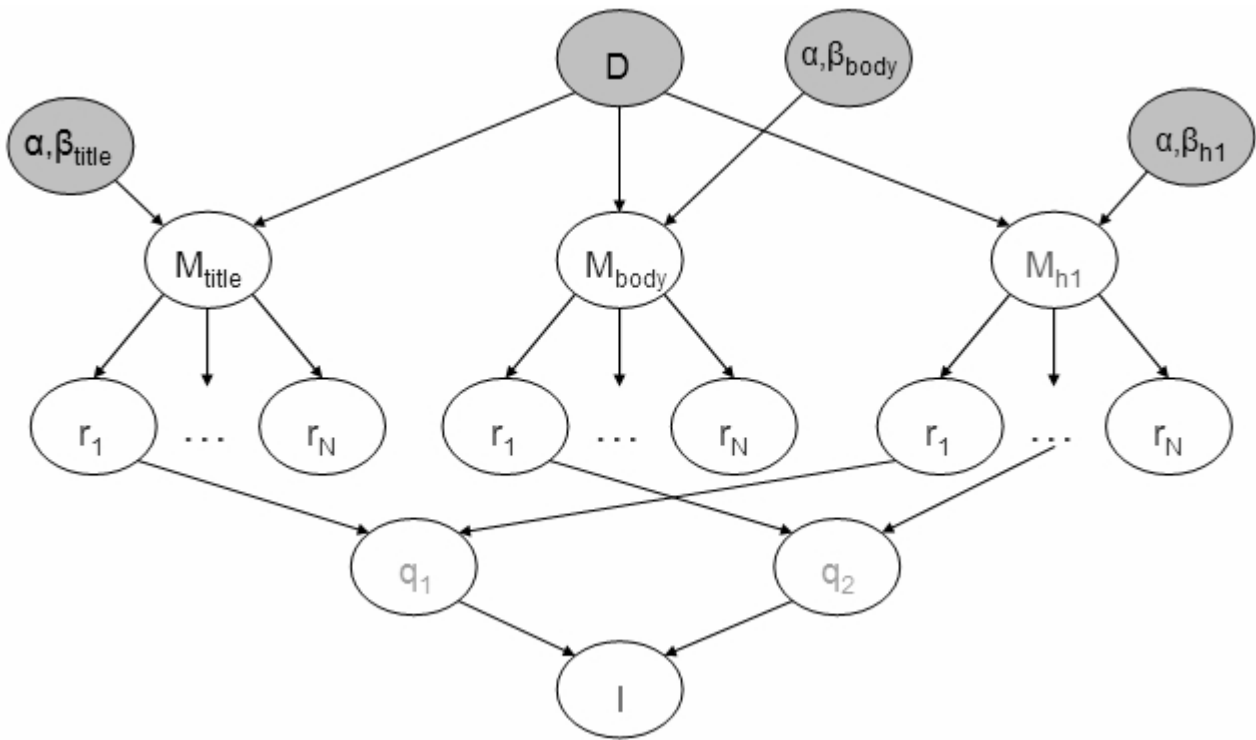
Hakujärjestelmänä tutkimuksessa käytetään Indri -hakukonetta, joka on osa Massachusettsin yliopiston kieliteknologian instituutin ja Carnegie Mellon -yliopiston Lemur-projektia. Indri hyödyntää kielimalliin perustuvaa todennäköisyyden laskentaa ja päättelyverkkoa (inference network).

Kielimallissa jokaiselle tietokannan dokumentille arvioidaan tilastollinen malli. Tämän jälkeen dokumentit pisteytetään sen mukaan, kuinka todennäköisesti dokumenttien kielimallista  $D$  parametrin arvoilla  $\alpha$  ja  $\beta$  voidaan johtaa tiedontarve  $I$ . Tämä voidaan esittää muodossa  $P(I|D, \alpha, \beta)$ . Tässä käytetyssä kielimallissa dokumenttia  $D$  kuvataan binääristen piirrevektoreiden ( $r$ ) joukkoina (multisets of binary feature vectors). Piirrevektori voi olla lähes mikä tahansa dokumenttia kuvaava piirre, kuten esimerkiksi sanan esiintyminen dokumentissa tai sanan sijainti lauseessa. Binäärisyys viittaa siihen, esiintyykö piirre dokumentissa vai ei.

Jokaiselle dokumentille arvioidaan moni-Bernoulli -malli ja lisäksi Indri arvioi dokumentin eri osille, kuten tiivistelmille ja kappaleille oman mallinsa. Näitä käsitellään pseudodokumentteina ja niitä voidaan hakea aivan kuten tavallisia dokumentteja.

Päättelyverkko muodostaa Indrin perustan. Päättelyverkon avulla voidaan yhdistää eri lähteistä saatua informaatiota siitä, onko dokumentti relevantti. Päättelyverkko koostuu kyselyverkosta  $I$  ja dokumenttiverkosta  $D$ . Päättelyverkossa oletetaan, että kysely koostuu käsitteistä (concepts), jotka voivat olla esimerkiksi termejä tai fraaseja. Oletuksena on, että dokumentti on käyttäjälle relevantti, jos se sisältää kyselyn käsitteet. Tässä pitää muistaa, että termin (esimerkiksi terrorismi) esiintyminen dokumentissa ei vielä takaa, että dokumentti käsittelee terrorismia. Päättelyverkon tehtävänä on arvioida, kuinka hyvin termin esiintyminen dokumentissa ennustaa dokumentin relevanssia. (Strohman, Metzler, Turtle, ja Croft 2005.)

**Kuva 5. Esimerkki päättelyverkosta (Metzler 2005)**



Kuvassa 4:

D = dokumentti

M viittaa kielimalleihin, jotka ovat mallintavat dokumentin representaatioita. Kuvan 4 dokumentilla on kolme kielimallia, jotka mallintavat otsikkoa (title), leipätekstiä (body) ja väliotsikoita (h).

$\alpha$  ja  $\beta$  ovat tasoitusparametreja (smoothing parametres), joiden avulla mm. estetään todennäköisyyksiä saamasta arvoa nolla.

r = representaatiosolmu (representation node), joka viittaa dokumentin piirteisiin, joita voidaan kuvata Indrin rakenteisilla kyselyillä

I = tiedontarve

q = kyselysolmu, joka määrittelee, miten dokumenttia kuvaavia piirteitä yhdistellään.

Malli olettaa että kaikki piirteet esiintyvät dokumentissa toisistaan riippumattomina, eli jonkun piirteen esiintyminen dokumentissa ei lisää todennäköisyyttä sille, että joku muu piirre esiintyy dokumentissa. Piirrevektorin ( $i$ ) todennäköisyys dokumentissa ( $D$ ) voidaan laskea kaavalla 16.

Kaava 16

$$P(r_i | D, \alpha_i, \beta_i) = \frac{\#(r_i, D) + \alpha_i}{|D| + \alpha_i + \beta_i}$$

Kun valitaan  $\alpha_i = \mu P(r_i | C)$  ja  $\beta_i = \mu(1 - P(r_i | C))$ , saadaan monijäseninen Bertoulli -malli, joka vastaa Dirilecht -tasoitusta. Kaavassa 17  $\mu$  on säädeltävä tasoitusparametri. (Strohman et al. 2005.)

Kaava 17

$$P(r_i | D, \alpha_i, \beta_i) = \frac{\#(r_i, D) + \mu P(r_i | C)_i}{|D| + \mu}$$

Prosessin tarkoituksena on laskea todennäköisyys sille, että dokumentti on relevantti kyselyyn. Kuvassa 4 todennäköisyys sille, että dokumentti D on relevantti, löytyy solmusta I. Solmun I ja representaatiolosojen r välissä on joukko kyselysolmuja q, jotka määrittävät sen, kuinka dokumentin relevanssin todennäköisyyden pitäisi riippua dokumenttiresentaatiosta. Kyselysolmut arvioivat representaatiolosuja sen mukaan, kuinka todennäköisesti käsite on ilmaistu dokumentissa. Kyselysolmujen käyttäytymisen, eli sen, miten representaatiolosmun arvioidaan kuvastavan dokumentin relevanssia, määrittää Indrin kyselykieli. (Strohman et al. 2005.) Indrin kyselykielestä voi katsoa lisää esimerkiksi edellä mainitusta Strohmanin artikkelista. Tässä tutkimuksessa kyselytermien yhdistämiseen käytetään Indrissä oletusarvoisesti käytettyä combine - operaattoria. Matemaattisesti combine vastaa InQuery -järjestelmän #and -operaattoria.

## **8.4 Kyselyjen suunnittelu**

Lähtötilanteena tutkimuksessa on kolmella eri tasolla onnistuneet kyselyt, joita kutsutaan tässä heikosti, hieman paremmin ja hyvin onnistuneiksi aloituskyselyiksi. Oleellista tässä on, että kyselyt ovat eritasoisesti onnistuneita, ei se, ovatko kyselyt hyviä tai huonoja sinänsä. Kokeellisissa tiedonhaku tutkimuksissa kyselyt muodostetaan tavallisesti automaattisesti hakuaiheiden otsikoiden tai kuvailukentissä esiintyvien sanojen perusteella. Tällöin kyselyistä tulee melko pitkiä, vaikka käyttäjien tekemät kyselyt ovat tunnetusti vain muutaman sanan mittaisia. Lisäksi automaattisesti muodostetuissa kyselyissä ei voida kontrolloida kyselyn onnistuneisuutta, vaikka käyttäjien tekemissä kyselyissä samasta hakuaiheesta esiintyy variaatiota onnistuneisuuden suhteen. Eritasoisten lyhyiden kyselyiden avulla voidaan näin ollen tuoda kokeelliseen tutkimukseen realismia, vaikka kyselyt ovatkin tutkimusta varten suunniteltuja. Eritasoisesti onnistuneita kyselyitä ei ollut saatavissa luonnollisista lähteistä, joten muodostin kyselyt tiedonhakupelin avulla tätä tarkoitusta varten.

Tiedonhakupeli eli QPA (= Query Performance Analyzer) on Tampereen yliopiston informaatiotutkimuksen laitoksen kehittämä web-sovellus, jota voidaan käyttää tiedonhaun ilmiöiden havainnollistamiseen. Tiedonhakupelissä valitaan ensin hakuaihe, dokumenttikokoelma ja tiedonhakujärjestelmä. Sitten käyttäjä syöttää kyselyn ja järjestelmä antaa tulosjoukon lisäksi palautteena saanti-tarkkuus -käyrän. Käyttäjä saa näin ollen tiedon myös niistä relevanteista dokumenteista, joita ei osu tulosjoukkoon. (Ks. myös Sormunen 2002b.)

Muodostin kyselyt eritasoisiksi kokeilemalla erilaisia hakuavainyhdistelmiä tiedonhakupelin antaman palautteen perusteella. Jokaiseen hakuaiheeseen, joita oli yhteensä 35, muodostin kolme eritasoista kyselyä. Aluksi tavoitteenani oli käyttää vain hakuaiheiden otsikko- ja kuvailukenttien sanoja. Tämä osoittautui kuitenkin mahdottomaksi ja otin kyselyihin mukaan myös relevanteissa dokumenteissa esiintyviä termejä. Jotta kyselyistä olisi saatu mahdollisimman realistisia, rajasin niiden pituuden maksimissaan kolmeen sanaan.

Kyselyn onnistuneisuutta mitattiin neliportaista relevanssia hyödyntäen siten, että jokaisesta tulosjoukossa esiintyvistä erittäin relevantista dokumentista kysely sai kolme pistettä, relevantista dokumentista kaksi ja marginaalisesti relevantista dokumentista yhden pisteen. Kysely sai pisteitä vain, mikäli kymmenen ensimmäisen dokumentin joukossa esiintyi edellä mainittuja dokumentteja. Kysely arvioitiin heikosti onnistuneeksi, mikäli tulosjoukon pisteet jäivät 1-5:een, kohtalaisen onnistuneeksi, mikäli se sai 5-9 pistettä ja hyvin onnistuneeksi, mikäli se sai 9-15 pistettä. Pistemäärä vaihteli sen mukaan, kuinka paljon aiheeseen oli relevantteja dokumentteja tietokannassa. Alla olevasta taulukosta 1. nähdään, että heikkotasaisen kyselyn pisteet ovat keskimäärin 3,11, keskitasoisen kyselyn pisteet 7,8 ja hyvätasoisen kyselyn pisteet 12,49.

**Taulukko 1. Eritasoisten kyselyjen pisteytys**

| Kyselyn onnistuneisuus | Relevanttien dokumenttien määrä liberaaleilla relevanssi-kriteereillä (1) | Relevanttien dokumenttien määrä keskitiukoilla relevanssi-kriteereillä (2) | Relevanttien dokumenttien määrä tiukoilla relevanssi-kriteereillä (3) | Kyselyn pisteet | Kyselyn pituus |
|------------------------|---|--|---|-----------------|----------------|
| Heikkotasoinen         | 1,4 (1-3)   | 1 (0-2)  | 0,7 (0-1)   | 3,1             | 2,4            |
| Keskitasoinen          | 3,6 (3-6)   | 2,6 (1-3)  | 1,6 (0-3)   | 7,8             | 2,5            |
| Hyvätasoinen           | 5,6 (4-8)   | 4,3 (2-6)  | 2,6 (1-5)   | 12,5            | 2,7            |

Jos esimerkiksi haku tuottaa kymmenen dokumentin kärkeen kaksi erittäin relevanttia dokumenttia yhden relevantin dokumentin ja yhden marginaalisesti relevantin dokumentin, on kyselyn saama pistemäärä  $2 \cdot 3 + 2 + 1 = 9$  pistettä. Ensimmäisen tulossivun jälkeisistä dokumenteista kysely ei saa pisteitä.

Tässä käytetty dokumenttien pisteytys perustuu intuitioon, eikä tarkoita sitä, että todellisuudessa yksi erittäin relevantti dokumentti vastaisi kolmea marginaalisesti relevanttia dokumenttia, vaikka periaatteessa näin voisi ollakin. Tämän tutkimuksen kannalta oleellista on eritasoisten kyselyiden muodostaminen ja tähän riittää yksinkertainen 0-3 pisteytys.

Taulukossa 1 on lisäksi esitetty, montako eritasoisesti relevanttia dokumenttia eritasoisesti onnistuneet kyselyt tuottavat eri tasoilla. Esimerkiksi heikkotasoisessa kyselyssä on keskimäärin 1,4 dokumenttia liberaaleilla relevanssikriteereillä mitattuna vaihteluvälin ollessa 1-3 dokumenttia. Relevanttien dokumenttien määrä tulosoukossa rajoittaa palautteeseen valittavien dokumenttien lukumäärää siten, että heikkotasoisesta kyselystä ei voida antaa palautetta kolmea dokumenttia enempää. Tämän seurauksena tulostaulukoissa esitetty viiden dokumentin relevanssipalautte tarkoittaa heikkotasoisesta kyselyn tapauksessa enintään kolmea dokumenttia ja kymmenestä dokumentista annettu palautte on identtinen viiden dokumentin palautteen kanssa. Sen sijaan hyvin onnistuneilla kyselyillä palautteen määrää liberaaleilla kriteereillä voidaan kasvattaa joidenkin kyselyiden tapauksessa aina kahdeksaan dokumenttiin asti. Mikäli tulosjoukossa ei ole relevantteja dokumentteja tiukoilla tai keskitiukoilla relevanssikriteereillä, palautetta ei anneta lainkaan eikä laajennusta tehdä. Tällöin laajennetulle kyselylle laskettu tuloksellisuus on identtinen laajentamattoman kyselyn tuloksellisuuden kanssa.

Tutkimuksessa ei käytetä oikeita käyttäjiä, mutta koeasetelmaan pyritään tuomaan realismia mallintamalla käyttäjien toimintaa. Neliportaisesti relevanssiarvioitujen dokumenttien avulla käyttäjätilannetta voidaan ensinnäkin simuloida relevanssikriteerien mukaan. Esimerkiksi liberaaleilla relevanssikriteereillä toimivalle käyttäjälle kelpaa marginaalisesti relevantit ja tätä relevantimmat dokumentit, kun taas tiukoilla relevanssikriteereillä toimiva käyttäjä kelpuuttaa mukaan vain erittäin relevantit dokumentit. Toisekseen käyttäjien toimintaa mallinnetaan sen mukaan, kuinka monesta dokumentista he ovat valmiita antamaan palautetta. Tämän tutkimuksen käyttäjäskenaariossa palautteeksi valitaan yksi, viisi tai kymmenen dokumenttia. Kolmanneksi käyttäjätilanteena voidaan nähdä kolmella eri tasolla onnistuneet kyselyt.

## **8.5 Tulosten analysointi ja evaluointi**

Tulosten evaluointiin käytetään keskitarkkuuksien keskiarvoa (MAP) ja tilastollisen merkitsevyyden testaamiseen Friedmanin kaksisuuntaista varianssianalyysiä. Tulosten käytännön merkittävyyden mittaamiseen käytetään Sparck-Jonesin (1974) käytännön merkittävyyttä mittaavaa sääntöä

### **8.5.1 Hakutehokkuuden evaluointi**

Tiedonhakatutkimuksessa tyypillinen hakutehokkuuden evaluointi perustuu saantiin ja tarkkuuteen. Saannilla tarkoitetaan löydettyjen relevanttien dokumenttien suhdetta kaikkiin relevantteihin dokumentteihin ja tarkkuudella löydettyjen relevanttien dokumenttien suhdetta kaikkiin löydettyihin dokumentteihin. Tässä tutkimuksessa hakutehokkuuden mittarina käytetään keskitarkkuuksien keskiarvoa (MAP = mean average precision), joka on yksi käytetyimmistä evaluointimittareista tiedonhakatutkimuksessa (Manning, Raghavan ja Schütze 2008, 147-148). Menetelmän ajatuksena on laskea tarkkuus jokaisen löydetyn relevantin dokumentin jälkeen ja laskea tarkkuuksien keskiarvo. MAP saadaan lasketuksi kullekin menetelmälle, kun lasketaan keskiarvo kaikkien kyselyiden keskitarkkuuksista. Tässä tutkimuksessa tulosjoukon koko on rajattu 100 dokumenttiin. Sellaiset relevantit dokumentit, jotka jäävät 100 dokumentin ulkopuolelle saavat tarkkuudeksi arvon 0.

Mitä asioita tulosten tulkinnassa huomioida, kun hakujärjestelmän tehokkuutta mitataan MAP:lla? MAP:lla lasketut tarkkuudet vaihtelevat suuresti riippuen siitä, kuinka monta relevanttia dokumenttia kokoelmassa on hakuaihetta kohden. Vaikka tulosjoukon kärjessä olisikin paljon relevantteja dokumentteja, saattaa MAP:lla laskettu hakutehokkuus jäädä pienehköksi, mikäli hakuaiheeseen on paljon relevantteja dokumentteja myös sadan kärkidokumentin ulkopuolella. Toisaalta, jos hakuaiheeseen on vain vähän relevantteja dokumentteja tietokannassa, mutta dokumentit löytyvät tulosjoukon kärkipäästä, on MAP:lla arvioitu hakutehokkuus korkea. Tämän vuoksi hakuaiheita pitäisi olla riittävästi ja niiden tulisi olla riittävän erilaisia, jotta tiedonhakujärjestelmän tehokkuudesta saadaan luotettavaa tietoa. MAP:n etuna on kuitenkin sen stabiilisuus ja hyvä erottelukyky. Lisäksi se on vakiinnuttanut paikkansa tiedonhakatutkimuksissa, mikä parantaa tulosten vertailtavuutta muihin tutkimuksiin. (Manning, Raghavan ja Schütze 2008, 147-148.) Tästä syystä myös tässä tutkimuksessa päädyttiin MAP:n käyttöön.

## 8.5.2 Tilastolliset menetelmät

Tulosten merkitsevyyden testaamiseen tarvitaan tilastollisia menetelmiä. Tässä tutkimuksessa tilastollisen merkitsevyyden testaamiseen käytetään Conoverin versiota Friedmanin kaksisuuntaisesta varianssinalyysistä (two-way analysis of variance) , sillä Conoverin versio on osoittautunut tehokkaaksi tiedonhaku tutkimuksessa. Friedmanin testi on epäparametrinen, mikä tarkoittaa sitä, että otosten ei tarvitse noudattaa normaalijakaumaa. Epäparametriset testit sopivat hyvin tiedonhaku tutkimukseen, sillä saanti ja tarkkuus ovat epäjatkuvia mittareita. Lisäksi testin avulla voidaan vertailla useita (yli kaksi) otoksia. Friedmanin testissä data esitetään riveinä  $b$  ja sarakkeina  $k$ , jossa riveillä esitetään tarkkuudet eri menetelmillä. Sitten tarkkuusluvut muunnetaan järjestysluvuiksi välille  $1-k$ . Friedmanin testi edellyttää, että muuttujat ovat vähintään järjestysasteikkolaisia. Se ei ota huomioon, kuinka paljon tarkkuudet vaihtelevat riveittäin. Lopuksi jokaisen menetelmän saamat järjestysluvut lasketaan yhteen ja Friedmanin testin avulla lasketaan, millä todennäköisyydellä muuttujien  $k$  yhteenlasketut järjestysluvut eroavat sattumalta saaduista arvoista. (ks. esim. Kekäläinen 1999, 80-82.) Testin ( $F$ ) laskemiseksi käytetään tässä seuraavaa kaavaa, joka on peräisin Conoverilta (1980, 299-300):

Kaava 18 
$$F = \frac{(b-1)[B - bk(k+1)^2 / 4]}{A - B}$$

jossa

Kaava 19 
$$A = \sum_{i=1}^b \sum_{j=1}^k [R(X_{ij})]^2$$

ja

Kaava 19 
$$B = \frac{1}{b} \sum_{j=1}^k R_j^2$$

$b$  = rivien lukumäärä

$k$  = sarakkeiden lukumäärä

$R(X_{ij})$  = solun järjestysluku rivillä  $i$  sarakkeessa  $j$

$R_j$  = järjestyslukujen summa sarakkeessa  $j$ . (Conover 1980, 299-300; ks. myös Kekäläinen 1999, 80-81.)

Tilastollinen testaus perustuu nollihypoteesiin, jonka mukaan menetelmien väriset erot ovat sattumaa. Nollihypoteesi hylätään, jos ainakin yksi testattavista menetelmistä eroaa ainakin yhdestä toisesta testattavasta menetelmästä. Näin tapahtuu, mikäli  $F$ :n arvo osoittautuu merkitseväksi.

Merkitsevyystaso määrää, millä todennäköisyydellä nollihypoteesi voidaan hylätä. Yleisesti käytettäviä merkitsevyystasoja ovat 0,05 (melkein merkitsevä), 0,01 (merkitsevä) ja 0,001 (erittäin merkitsevä). Esimerkiksi taso 0,05 tarkoittaa, että 5% tapauksista ero nollihypoteesiin johtuu sattumasta. Näitä merkitsevyystasoja käytetään myös tässä tutkimuksessa. Edellä esitetyt kaavat eivät kuitenkaan kerro, mitkä vertailtavista menetelmistä eroavat toisistaan. Siksi jokaista menetelmää on vertailtava pareittain seuraavan kaavan mukaisesti:

Kaava 20 
$$|R_j - R_i| \geq t - \alpha / 2 \left[ \frac{2b(A - B)}{(b - 1)(k - 1)} \right]^{1/2}$$

jossa

$R_j$  ja  $R_i$  ovat sarakkeiden  $j$  ja  $i$  järjestyslukujen summa

$t - \alpha / 2$  antaa raja-arvon jonka yläpuolella on  $1 - \alpha / 2$  prosenttia jakaumasta

$\alpha$  on vakio, jonka arvo saadaan testiin liittyvistä liitetaulukoista. (Conover 1980, 299-300; ks. myös Kekäläinen 1999, 80-81; Siegel ja Castellan 1988, 8-9 .)

Vaikka kahden hakutavan mitattu tehokkuuden ero olisi tilastollisesti merkitsevä, ei erolla välttämättä ole käytännön merkitystä. Erojen käytännöllisen merkityksen mittaamiseen Sparck-Jones (1974) on kehittänyt tulosten käytännön merkittävyyttä mittaavan säännön, jonka mukaan vertailtavien menetelmien väliset erot ovat vähäisiä, mikäli ero on alle 5 prosenttiyksikköä, merkittäviä, jos ero on 5-10 prosenttiyksikköä ja olennaisia, mikäli ero on yli 10 prosenttiyksikköä. Tätä tulosten käytännön merkittävyyssääntöä hyödynnetään tässä tutkimuksessa tilastollisesti merkitsevien tulosten tulkinnessa.

## 8.6 Esitutkimus

Esitutkimuksessa tarkoituksenani oli selvittää, kuinka monta laajennustermiä RATF -kaavalla tulee valita relevanssipalautteeseen optimaalisen hakutehokkuuden aikaansaamiseksi.

Lähtökohtana oli, että jokaisesta relevanssipalautteeseen valitusta dokumentista poimitaan 50 suurimman RATF -arvon saanutta termiä. Tämän jälkeen dokumenttien top 50 -listoista muodostetaan uusi lista siten, että kärkeen sijoitetaan sellaiset termit, jotka esiintyvät mahdollisimman monen dokumentin top 50 -listassa mahdollisimman suurella RATF -arvolla. Jos



esimerkiksi ”käyttäjä” poimii relevanssipalautteeseen neljä dokumenttia, niin uuden top -listan kärkeen sijoittuvat sellaiset termit, jotka esiintyvät kaikissa neljässä dokumentissa riittävän suurella RATF -arvolla, seuraavana listassa ovat kolmessa dokumentissa esiintyneet termit jne. Olen valinnut listasta 10, 20, 30 ja 50 ensimmäistä sanaa laajennettuun hakuun ja kokeillut, kuinka laajennus vaikuttaa haun tuloksellisuuteen. Tähän tarkoitukseen olen valinnut kahdeksan keskitasoisesti onnistunutta kyselyä TRECUTA dokumenttikokoelmasta.

Laajennustermien määrän vertailua varten olen hyödyntänyt luvussa 8.2 esitettyä kyselyiden pisteytystä. Keskitasolla onnistuneiden vertailukyselyiden saama pistemäärä on keskimäärin 6,3 (luku on keskiarvo kahdeksan keskitasoisesti onnistuneen kyselyn pisteistä, eikä vastaa taulukossa 1 esitettyä keskitasoisen kyselyn pistemäärää). Kun relevanssipalautteella tehtyyn laajennukseen otetaan mukaan 10 hakutermiä, nousee kyselyn pistemäärä keskimäärin 11,5:een käyttäjän preferensseistä riippumatta. Kun mukaan otetaan 20 laajennustermiä, nousee laajennettujen kyselyiden pistekeskisarvo 12,4:ään. 30 laajennustermillä kyselyiden pistekeskisarvo nousee edelleen 13:een. 50 laajennustermillä kyselyiden pistekeskisarvoksi tulee 12,6 Keskihajonta kaikissa laajennuksissa on n. 5,5, eli vaihtelu kyselyiden (tai hakuaiheiden) välillä on melko runsasta.

Tässä kohtaa on huomioitava se, että laajennettun kyselyn tuloslistassa on mukana samoja dokumentteja kuin laajentamattoman kyselyn listassa. Tämä tarkoittaa sitä, että käyttäjän jo löytämät relevantit dokumentit on laskettu relevanteiksi myös uudessa tuloslistassa, vaikka ne eivät käyttäjälle uusia enää olekaan. Tuloksellisuuserot laajennustermien määrän suhteen ovat melko pieniä, mutta näiden suuntaa antavien tulosten perusteella päätin kuitenkin tehdä laajennuksen 30 termillä. Tämä tuotti parhaan hakutehokkuuden (13 pistettä) pistekeskisarvolla mitattuna.

## **8.7 Lopullinen koeasetelma**

Koe toteutettiin seuraavasti:

1. Muokataan eritasoisesti onnistuneet kyselyt tiedonhakupelin avulla
2. Ajetaan kyselyt Lemur Indri -hakujärjestelmässä, jossa kyselytermit on liitetty toisiinsa oletusarvoisella combine -operaattorilla.
3. Poimitaan relevanssipalautteeseen enintään 1, 5 tai 10 relevanttia dokumenttia eri relevanssikriteereillä 10 ensimmäisen dokumentin joukosta. Dokumenttien määrä riippuu saatavuudesta, sillä 10 ensimmäisen dokumentin joukossa ei käytännössä ole 10 relevanttia

dokumenttia. Pseudorelevanssipalautteessa tuloslistan kärjestä valitaan 1, 5 tai 10 dokumenttia riippumatta siitä, ovatko ne relevantteja vai eivät.

4. Jokaisesta relevanssipalautteeseen valitusta dokumentista (joko ”käyttäjän” valitsemista tai automaattisesti valituista) poimitaan 50 suurimman RATF -arvon saanutta termiä. Ennen tätä stop-sanat on kuitenkin poistettu ja termit on perusmuotoistettu. Tämän jälkeen dokumenttien top 50 -listoista poimitaan termejä ja muodostetaan uusi lista siten, että kärkeen sijoitetaan sellaiset termit, jotka esiintyvät mahdollisimman monen dokumentin top 50 -listassa mahdollisimman suurella RATF -arvolla. Jos esimerkiksi ”käyttäjä” poimii relevanssipalautteeseen neljä dokumenttia, niin uuden top -listan kärkeen sijoittuvat sellaiset termit, jotka esiintyvät kaikissa neljässä dokumentissa riittävän suurella RATF -arvolla, seuraavana listassa ovat kolmessa dokumentissa esiintyneet termit jne. Näin muodostetusta listasta valitaan 30 ensimmäistä termiä. 30:n termiin päädyttiin esitutkimuksen perusteella.
5. Lisätään laajennustermit ( $l$ ) alkuperäiseen kyselyyn ( $a$ ) Indrin oletusarvoisella combine - operaattorilla. Näin laajennettu kysely noudattaa rakennetta  
$$\#combine (termi_{a_1}, termi_{a_2}, termi_{a_3}, \dots, termi_{l_1}, termi_{l_2}, termi_{l_3}, \dots).$$
6. Relevanssipalautteen tuloksellisuuden laskentaan käytetään luvussa 6 esiteltyä jäädytysmenetelmää, jossa uusi laajennettu kysely ajetaan tietokantaan, josta on poistettu jo nähdyt 10 dokumenttia, joista relevanssipalautte on annettu. Tämän jälkeen jäljelle jääneet dokumentit järjestetään uudelleen laajennetulla kyselyllä. Laajennetun kyselyn tuloksellisuutta verrataan tilanteeseen, jossa käyttäjä ei anna relevanssipalautetta, vaan jatkaa tulosjoukon selaamista. Pseudorelevanssipalautteessa jäädytystä ei tarvitse tehdä, sillä ”käyttäjä” saa nähtäväkseen vain yhden tuloslistan, eikä jo nähtyjen dokumenttien ongelmaa ole.
7. Hakutehokkuutta evaluoidaan kyselyiden keskitarkkuuksien keskiarvolla (MAP). MAP lasketaan 100 kärkidokumentin joukosta. Joukon ulkopuolelle jääneet relevantit dokumentit saavat tarkkudeksi arvon 0.

## 9. Tulokset

Tulosten evaluointiin käytetään tässä tutkimuksessa keskitarkkuuksien keskiarvoa (MAP = mean average precision). Laajennettujen kyselyiden keskitarkkuutta verrataan laajentamattomien kyselyiden keskitarkkuuteen ja lisäksi eri laajennuksia verrataan keskenään.

Laajentamattomat, eritasoisesti onnistuneet kyselyt on kuitenkin muodostettu taulukon 1 pisteytyksen avulla kymmenen kärkidokumentin joukosta. Varsinaisena evaluointimittarina käytetty MAP sen sijaan on keskitarkkuuksien keskiarvo sadan dokumentin joukosta. Jos esimerkiksi kysely onnistuu heikosti pisteytyksellä arvioituna, onko sama kysely edelleen heikosti onnistunut, kun evaluointimittarina käytetään MAP:ia? Voihan olla, että kymmenen kärkidokumentin jälkeen löytyy runsaasti relevantteja dokumentteja, jolloin alun perin heikkotasaisen kyselyn onnistuneisuus muuttuukin MAP:lla laskettuna hyväksi. Toisin sanoen, miten kymmenen kärkidokumentin perusteella pisteytetyt kyselyt ennustavat kyselyn onnistuneisuutta, kun evaluointimittarina käytetään MAP:ia? Taulukosta 2 nähdään, että pisteytyksen avulla tuotetut tasoerot kyselyissä näkyvät selvästi myös MAP:lla lasketuissa tuloksissa. Pisteytyksellä heikoksi arvioitua kysely pysyy heikkotasaisena myös MAP:lla laskettuna. Tältä osin pisteytysmenetelmää voidaan pitää onnistuneena.

Mielenkiintoinen ero pisteytykseen verrattuna on, että taulukossa 2 relevanssikriteereiden tiukentuessa tasolta 1 tasolle 3, myös MAP:lla laskettu hakutehokkuus paranee. Pisteytyksessä (taulukko 1) tapahtuu käännteinen ilmiö, eli mitä tiukemmat relevanssikriteerit, sitä vähemmän relevantteja dokumentteja löytyy kymmenen kärkidokumentin joukosta. Keskitarkkuuden paraneminen relevanssikriteerien kiristyessä MAP:illa laskettuna johtuu siitä, että marginaalisesti relevantteja dokumentteja on enemmän dokumenttikannassa kuin erittäin relevantteja dokumentteja. Mitä enemmän hakuaiheeseen on relevantteja dokumentteja, sitä enemmän niitä jää myös 100 ensimmäisen dokumentin ulkopuolelle ja tämä laskee keskitarkkuutta.

Taulukko 2. **Eritasoisten laajentamattomien kyselyiden tuloksellisuus (MAP)**

| Kyselyn onnistuneisuus | 1    | 2    | 3    |
|------------------------|------|------|------|
| Heikko                 | 3,5  | 3,7  | 6,4  |
| Keskitaso              | 10,6 | 13   | 13,7 |
| Hyvä                   | 18,6 | 22,5 | 27   |

Varsinaiset tutkimustulokset on järjestetty taulukoihin 3-5 ja lukuihin 9.1 - 9.3 sen mukaan, millä relevanssikriteereillä tuloksellisuus on laskettu. Luvussa 9.1 marginaalisestikin relevantit dokumentit on laskettu relevanteiksi hakutehokkuutta laskettaessa. Luvussa 9.2 vain relevantit ja erittäin relevantit dokumentit huomioidaan ja niin edelleen. Taulukoiden sisällä tulokset on järjestetty edelleen laajentamattoman kyselyn onnistuneisuuden mukaan. Esimerkiksi heikkotasaisen aloituskyselyn laajennuksia verrataan heikkotasaiseen aloituskyselyyn. Numerot 1, 5 ja 10 viittaavat siihen, kuinka monta dokumenttia palautteeseen on valittu enimmillään. Pseudorelevanssipalautteessa numerot ovat kiinteitä ja esimerkiksi palaute 10 dokumentista myös tarkoittaa palautetta 10 dokumentista.

Lisäksi taulukoiden tulokset on eritelty sen mukaan, mitä relevanssitasoa dokumentin tulee vähintään edustaa, jotta se tulee valituksi relevanssipalautteeseen. RF 1 viittaa siihen, että myös marginaalisesti relevanteista dokumenteista on annettu relevanssipalautetta. Tasolla RF 2 palautteeseen on laskettu vain relevantit ja erittäin relevantit dokumentit ja tasolla RF 3 palautetta on annettu vain erittäin relevanteista dokumenteista. Mikäli palautteeseen valitaan esimerkiksi vain yksi dokumentti, valitaan aina relevantin mahdollinen palautteenantovaiheen relevanssikriteereistä riippumatta. Pseudorelevanssipalautteessa palaute annetaan automaattisesti samoista dokumenteista ja mukana ovat niin epärelevantit kuin erittäin relevantit dokumentit.

Tulosten raportointi lukujen 9.1 - 9.3 sisällä jäsentyy siten, että ensin kerrotaan relevanssipalautteen vaikutus heikkotasaiseen kyselyyn. Tämän jälkeen verrataan heikkotasaiselle kyselylle toteutettuja relevanssipalautemenetelmiä keskenään. Eri relevanssipalautemenetelmien väliset erot prosenttiyksikköinä nähdään suoraan taulukoista 3 - 5, mutta vertailuiden tilastolliset merkitsevyydestit on esitetty liitteessä. Sitten siirrytään keskitasoisesti onnistuneeseen aloituskyselyyn, jonka jälkeen seuraa hyvin onnistunut kysely. Lopuksi raportoidaan pseudorelevanssipalautteen vaikutus hakutehokkuuteen ja vertailu muihin palautemenetelmiin.

## ***9.1 Tulokset liberaaleilla relevanssikriteereillä mitattuna***

Taulukosta 3 nähdään, että heikkotasaisesta kyselyä voidaan parantaa relevanssipalautteen avulla hieman alle kymmenen prosenttiyksikköä laajentamattomaan kyselyyn verrattuna. Mitattu taso on

tilastollisesti erittäin merkitsevä ja se on merkittävä myös Sparck-Jonesin (1974) käytännön merkittävyyttä mittaavan säännön mukaan. Eri relevanssipalautemenetelmien välillä ei ole tilastollisesti merkitseviä eroja (ks. liite, taulukko 1). Sen sijaan pseudorelevanssipalautteeseen verrattuna relevanssipalautteeseen toimii huomattavasti paremmin.

Keskitasoisesti onnistunut aloituskysely hyötyy relevanssipalautteesta kaikissa tapauksissa. Parhaimmillaan aloituskyselyä voidaan parantaa relevanssipalautteen avulla jopa 13 prosenttiyksikköä. Hakutehokkuus kasvaa tällöin tilastollisesti erittäin merkittävästi ja olennaisesti myös käytännön merkittävyyttä arvioivalla säännöllä. Relevanssipalautemenetelmien välisessä vertailussa liberaaleilla relevanssikriteereillä (RF 1) annettu palaute enintään viidestä dokumentista näyttäisi toimivan tiukkojen kriteerien (RF 3) palautetta paremmin ja Friedmanin testin mukaan erot palautemenetelmien välillä ovat tilastollisesti merkitseviä (ks. liite, taulukko 2). Käytännössä ero on kuitenkin melko pieni. Pelkkä relevanssikriteerien variointi palautteenantovaiheessa ei tuo esiin käytännön tasolla merkittäviä eroja eri relevanssipalautemenetelmien välille. Relevanssipalautteen määrän kasvattaminen yhdestä viiteen sen sijaan tuottaa käyttäjälle käytännössä merkittävän hyödyn.

**Taulukko 3. Keskitarkkuudet (MAP) kyselyn onnistuneisuuden, palautedokumenttien määrän (RF dok.) ja palautteenantokriteerien (1-3) mukaan. Tuloksellisuus laskettuna liberaaleilla relevanssikriteereillä.**

| RF dok.  | RF 1           | Erotus | RF 2           | Erotus | RF 3           | Erotus | PRF           | Erotus |
|--|----------------|--------|----------------|--------|----------------|--------|---------------|--------|
| Heikosti onnistunut aloituskysely. Vertailukysely <b>3,5</b> (MAP).        |                |        |                |        |                |        |               |        |
| 1  | <b>12,6***</b> | 9,1    | <b>11,6***</b> | 8,1    | <b>10,9***</b> | 7,4    | <b>3,2***</b> | -0,3   |
| 5  | <b>13,3***</b> | 9,8    | <b>11,9***</b> | 8,4    | <b>10,9***</b> | 7,4    | <b>4,6</b>    | 1,1    |
| 10   | <b>13,3***</b> | 9,8    | <b>11,9***</b> | 8,4    | <b>10,9***</b> | 7,4    | <b>4,9</b>    | 1,4    |
| Keskitasoisesti onnistunut aloituskysely. Vertailukysely <b>10,6</b> (MAP) |                |        |                |        |                |        |               |        |
| 1  | <b>18,5***</b> | 7,9    | <b>18,5***</b> | 7,9    | <b>18,3***</b> | 7,7    | <b>8,4</b>    | -2,2   |
| 5  | <b>23,6***</b> | 13     | <b>22,5***</b> | 11,9   | <b>20,1***</b> | 9,5    | <b>12,3</b>   | 1,7    |
| 10   | <b>23,7***</b> | 13,1   | <b>22,5***</b> | 11,9   | <b>20,1***</b> | 9,5    | <b>12,3</b>   | 1,7    |
| Hyvin onnistunut aloituskysely. Vertailukysely <b>18,6</b> (MAP)           |                |        |                |        |                |        |               |        |
| 1  | <b>29,5***</b> | 10,9   | <b>29,5***</b> | 10,9   | <b>29,5***</b> | 10,9   | <b>18,4</b>   | -0,2   |
| 5  | <b>34,7***</b> | 16,1   | <b>34,5***</b> | 15,9   | <b>32,4***</b> | 13,8   | <b>23,6</b>   | 5      |
| 10   | <b>35,2***</b> | 16,6   | <b>34,7***</b> | 16,1   | <b>32,4***</b> | 13,8   | <b>20,7</b>   | 2,1    |

*Selite: Taulukon 4 ensimmäinen sarake kertoo palautteeseen valittujen dokumenttien lukumäärän. Tummennettuna esitetään laajennetun kyselyn keskitarkkuus. Tähdet numeroiden perässä kertovat tilastollisen merkitsevyyden suhteessa vertailukyselyyn. RF 1, RF 2 ja RF 3 viittaavat relevanssipalautteessa käytettyjen dokumenttien vähimmäisrelevanssitason. PRF tarkoittaa pseudorelevanssipalautetta. Vertailukyselyjen keskitarkkuusluvut (MAP) ovat suoraan taulukosta 2.*

Hyvin onnistunutta kyselyä voidaan parantaa relevanssipalautteen avulla 10,9 - 16,6 prosenttiyksikköä laajentamattomaan kyselyyn verrattuna. Tulos on tilastollisesti erittäin merkitsevä ja myös käytännön merkittävyyden kannalta mukaan olennainen. Palautemenetelmien keskinäisessä vertailussa liberaaleilla relevanssikriteereillä (RF 1) annettu palaute näyttäisi toimivan hieman tiukimpien kriteerien (RF 3) palautetta paremmin, mutta palautedokumenttien relevanssitason väliset erot eivät ole tilastollisesti merkitseviä. Sen sijaan palautedokumenttien enimmäismäärän kasvattaminen yhdestä viiteen tuottaa yli viiden prosenttiyksikön eron liberaaleilla (RF 1) ja keskitiukoilla (RF 2) relevanssikriteereillä. Tulos on tilastollisesti erittäin merkitsevä ja myös käytännössä merkittävä (ks. liite, taulukko 3).

Pseudorelevanssipalautteen avulla ei voida parantaa hakutehokkuutta tilastollisesti merkitsevästi. Näin käy riippumatta kyselyn onnistuneisuudesta. Yhdestä dokumentista annettu pseudopalaute näyttäisi jopa heikentävän hakutehokkuutta kaiktasoisilla kyselyillä.

## ***9.2 Tulokset keskitiukoilla relevanssikriteereillä mitattuna***

Taulukosta 4 nähdään, että heikkotasosta kyselyä voidaan parantaa relevanssipalautteen avulla hieman alle kymmenen prosenttiyksikköä laajentamattomaan kyselyyn verrattuna. Ero laajentamattomaan kyselyyn on tilastollisesti erittäin merkitsevä ja se on merkittävä myös käytännön tasolla. Kun palautemenetelmiä verrataan keskenään, huomataan että palautteessa käytettyjen dokumenttien relevanssitason nostamisella ei ole suurta merkitystä relevanssipalautteen toimivuuteen. Myöskään relevanssipalautteen määrällä ei ole heikkotasoisien kyselytapauksissa vaikutusta palautteen toimivuuteen, sillä palautteenantovaiheen relevanssikriteerien tiukentaminen vähentää myös mahdollisten palautedokumenttien määrää heikkotasoisella kyselyllä (ks. liite, taulukko 4).

Keskitasoisesti onnistunutta aloituskyselyä voidaan parantaa relevanssipalautteen avulla parhaimmillaan 13,3 prosenttiyksikköä relevanssipalautteen avulla. Hakutehokkuuden kasvu on tällöin tilastollisesti erittäin merkitsevä ja myös käytännössä olennainen. Jo yhdestä dokumentista annettu relevanssipalaute nostaa hakutehokkuutta yli 8 prosenttiyksikköä. Kun eri relevanssipalautemenetelmiä verrataan keskenään, huomataan että palautteessa käytettyjen dokumenttien relevanssitasolla on vain vähäinen merkitys hakutehokkuuteen. Esimerkiksi

liberaaleilla relevanssikriteereillä (RF 1) annettu palaute viidestä dokumentista näyttäisi toimivan tiukkojen kriteerien (RF 3) palautetta paremmin ja Friedmanin testin mukaan erot näiden palautemenetelmien välillä ovat tilastollisesti merkitseviä, mutta erotus jää kuitenkin alle viiteen prosenttiyksikköön. Sen sijaan palautedokumenttien enimmäismäärän kasvattaminen yhdestä viiteen tuottaa viiden prosenttiyksikön eron, kun palautteeseen valitaan dokumentteja liberaaleilla relevanssikriteereillä. Tulos on tilastollisesti erittäin merkitsevä ja myös käytännössä merkittävä (ks. liite, taulukko 5). Keskitiukoilla palautekriteereillä päästään myös lähes viiden prosenttiyksikön parannuksiin, kun palautedokumenttien määrää kasvatetaan yhdestä viiteen, mutta tulokset eivät ole tilastollisesti merkitseviä.

**Taulukko 4. Keskitarkkuudet (MAP) kyselyn onnistuneisuuden, palautedokumenttien määrän (RF dok.) ja palautteenantokriteerien (1-3) mukaan. Tuloksellisuus laskettuna liberaaleilla relevanssikriteereillä.**

| RF dok.  | RF 1           | Erotus | RF 2           | Erotus | RF 3           | Erotus | PRF           | Erotus |
|--|----------------|--------|----------------|--------|----------------|--------|---------------|--------|
| Heikosti onnistunut aloituskysely. Vertailukysely <b>3,7</b> (MAP)       |                |        |                |        |                |        |               |        |
| 1  | <b>13,4***</b> | 9,7    | <b>12,7***</b> | 9      | <b>12***</b>   | 8,3    | <b>3,1***</b> | -0,6   |
| 5  | <b>12,6***</b> | 8,9    | <b>13,1***</b> | 9,4    | <b>12***</b>   | 8,3    | <b>3,9</b>    | 0,2    |
| 10   | <b>12,6***</b> | 8,9    | <b>13,1***</b> | 9,4    | <b>12***</b>   | 8,3    | <b>4,5</b>    | 0,8    |
| Keskitasoisesti onnistunut aloituskysely. Vertailukysely <b>13</b> (MAP) |                |        |                |        |                |        |               |        |
| 1  | <b>21,2***</b> | 8,2    | <b>21,2***</b> | 8,2    | <b>21,1***</b> | 8,1    | <b>10,5</b>   | -2,5   |
| 5  | <b>26,1***</b> | 13,1   | <b>25,9***</b> | 12,9   | <b>23,3***</b> | 10,3   | <b>13,9</b>   | 0,9    |
| 10   | <b>26,3***</b> | 13,3   | <b>25,9***</b> | 12,9   | <b>23,3***</b> | 10,3   | <b>15,2</b>   | 2,2    |
| Hyvin onnistunut aloituskysely. Vertailukysely <b>22,5</b> (MAP)         |                |        |                |        |                |        |               |        |
| 1  | <b>36,2***</b> | 13,7   | <b>36,2***</b> | 13,7   | <b>36,2***</b> | 13,7   | <b>20,3</b>   | -2,2   |
| 5  | <b>41***</b>   | 18,5   | <b>41,8***</b> | 19,3   | <b>39,1***</b> | 16,6   | <b>25,6</b>   | 3,1    |
| 10   | <b>40,9***</b> | 18,4   | <b>41,9***</b> | 19,4   | <b>39,1***</b> | 16,6   | <b>23,5</b>   | 1      |

*Selite: Taulukon 4 ensimmäinen sarake kertoo palautteeseen valittujen dokumenttien lukumäärän. Tummennettuna esitetään laajennetun kyselyn keskitarkkuus. Tähdet numeroiden perässä kertovat tilastollisen merkitsevyyden suhteessa vertailukyselyyn. RF 1, RF 2 ja RF 3 viittaavat relevanssipalautteessa käytettyjen dokumenttien vähimmäisrelevanssitason. PRF tarkoittaa pseudorelevanssipalautetta. Vertailukyselyjen keskitarkkuusluvut (MAP) ovat suoraan taulukosta 2.*

Hyvin onnistunutta aloituskyselyä voidaan parantaa relevanssipalautteen avulla 13,7 - 19,4 prosenttiyksikköä. Tulos on tilastollisesti erittäin merkitsevä ja myös käytännössä olennainen. Eri palautemenetelmiä vertailtaessa keskenään, nähdään että palautteessa käytettyjen dokumenttien relevanssitason välillä ei ole tilastollisesti merkitseviä eroja, mutta relevanssipalautteen määrän kasvattaminen yhdestä viiteen dokumenttiin tuottaa merkitsevän eron (ks. liite, taulukko 6).

Palautteen määrän kasvattamisella on myös käytännön merkittävyyttä mittaavan säännön mukaan merkitystä, mikäli käyttäjä toimii keskitiukoilla tai liberaaleilla relevanssikriteereillä.

Pseudorelevanssipalautteella ei voida vaikuttaa tilastollisesti merkitsevästi hakutehokkuuteen, eikä aloituskyselyn onnistuneisuudella ole vaikutusta pseudopalautteen toimivuuteen.

### **9.3 Tulokset tiukoilla relevanssikriteereillä mitattuna**

Taulukosta 5 nähdään, että heikkotasosta kyselyä voidaan parantaa relevanssipalautteen avulla yli 12 prosenttiyksikköä relevanssipalautteen avulla, kun tuloksellisuus lasketaan keskitiukoilla relevanssikriteereillä. Tulos on tilastollisesti erittäin merkitsevä ja myös käytännön merkittävyyttä mittaavan säännön mukaan olennainen. Eri palautemenetelmien väliset erot ovat hyvin pieniä, eivätkä yllä tilastolliseen merkitsevyyteen (ks. liite, taulukko 7).

Relevanssipalautteen vaikutus keskitasoisesti onnistuneeseen kyselyyn on hieman suurempi kuin heikkotasaisen kyselyn tapauksessa. Tiukoilla relevanssikriteereillä annetun palautteen avulla päästään jopa 14 prosenttiyksikköä parempaan hakutehokkuuteen kuin laajentamattomalla kyselyllä. Eri relevanssipalautemenetelmät tuottavat kuitenkin keskenään varsin samankaltaisia tuloksia, eikä edes palautedokumenttien määrällä ole vaikutusta relevanssipalautteen toimivuuteen.



**Taulukko 5. Keskitarkkuudet (MAP) kyselyn onnistuneisuuden, palautedokumenttien määrän (RF dok.) ja palautteenantokriteerien (1-3) mukaan. Tuloksellisuus laskettuna liberaaleilla relevanssikriteereillä.**

| RF dok.  | RF 1           | Erotus | RF 2           | Erotus | RF 3           | Erotus | PRF           | Erotus |
|--|----------------|--------|----------------|--------|----------------|--------|---------------|--------|
| Heikosti onnistunut aloituskysely. Vertailukysely <b>6,4</b> (MAP)         |                |        |                |        |                |        |               |        |
| 1  | <b>19***</b>   | 12,6   | <b>19***</b>   | 12,6   | <b>18,4***</b> | 12     | <b>4,5**</b>  | -1,9   |
| 5  | <b>16,6***</b> | 10,2   | <b>19,2***</b> | 12,8   | <b>18,4***</b> | 12     | <b>5,7</b>    | -0,7   |
| 10   | <b>16,6***</b> | 10,2   | <b>19,2***</b> | 12,8   | <b>18,4***</b> | 12     | <b>5,4</b>    | -1     |
| Keskitasoisesti onnistunut aloituskysely. Vertailukysely <b>13,7</b> (MAP) |                |        |                |        |                |        |               |        |
| 1  | <b>24,7***</b> | 11     | <b>24,7***</b> | 11     | <b>24,8***</b> | 11,1   | <b>11,9</b>   | -1,8   |
| 5  | <b>27***</b>   | 13,3   | <b>27***</b>   | 13,3   | <b>27,7***</b> | 14     | <b>16,5</b>   | 2,8    |
| 10   | <b>27,1***</b> | 13,4   | <b>27***</b>   | 13,3   | <b>27,7***</b> | 14     | <b>16,8**</b> | 3,1    |
| Hyvin onnistunut aloituskysely. Vertailukysely <b>27</b> (MAP)             |                |        |                |        |                |        |               |        |
| 1  | <b>46,8***</b> | 19,8   | <b>46,8***</b> | 19,8   | <b>46,8***</b> | 19,8   | <b>25,1</b>   | -1,9   |
| 5  | <b>47,6***</b> | 20,6   | <b>44,4***</b> | 17,4   | <b>50,6***</b> | 23,6   | <b>31,4</b>   | 4,4    |
| 10   | <b>43,7***</b> | 16,7   | <b>44,4***</b> | 17,4   | <b>50,6***</b> | 23,6   | <b>25,9</b>   | -1,1   |

*Selite: Taulukon 4 ensimmäinen sarake kertoo palautteeseen valittujen dokumenttien lukumäärän. Tummennettuna esitetään laajennetun kyselyn keskitarkkuus. Tähdet numeroiden perässä kertovat tilastollisen merkitsevyyden suhteessa vertailukyselyyn. RF 1, RF 2 ja RF 3 viittaavat relevanssipalautteessa käytettyjen dokumenttien vähimmäisrelevanssitason. PRF tarkoittaa pseudorelevanssipalautetta. Vertailukyselyjen keskitarkkuusluvut (MAP) ovat suoraan taulukosta 2.*

Relevanssipalautteen vaikutus hyvin onnistuneisiin aloituskyselyihin on huomattava kaikilla relevanssitasoilla ja palautemäärillä niin tilastollisesti kuin käytännössä. Tiukoilla relevanssikriteereillä viidestä dokumentista annettu palaute tuottaa jopa 23,6 prosenttiyksikköä paremman tuloksellisuuden laajentamattomaan kyselyyn verrattuna. Palautteessa käytettyjen dokumenttien relevanssitason nostamisella ei kuitenkaan ole tilastollisesti merkitsevää vaikutusta hakutehokkuuteen. Laajennukseen valittavien dokumenttien määrän kasvattaminen yhdestä viiteen tuottaa tilastollisesti merkitsevän eron, kun palautetta annetaan vain erittäin relevanteista dokumenteista, (3,8 prosenttiyksikköä), mutta tuloksella ei ole käytännön merkittävyyttä (ks. liite, taulukko 9).

Pseudorelevanssipalautteen avulla ei voida parantaa tilastollisesti merkitsevästi hakutehokkuutta. Erityisen heikkoja tulokset ovat heikkotasoisella kyselyllä, jolloin hakutehokkuus jopa heikkenee hieman. Keskitasoista kyselyä voidaan parantaa pseudorelevanssipalautteella kolmisen prosenttiyksikköä, kun palautetta annetaan kymmenestä dokumentista. Tulos on myös tilastollisesti merkitsevä, vaikka jääkin käytännön merkittävyydeltään vaatimattomaksi.

## **9.4 Yhteenvetoa tuloksista**

### 1. Käyttäjän relevanssipalautteen tuloksellisuus

Tämän tutkimuksen perusteella relevanssipalautteella voidaan parantaa kyselyiden hakutehokkuutta verrattuna tilanteeseen, jossa palautetta ei anneta. Parhaimmillaan ero on jopa 23,6 prosenttiyksikköä, kun käyttäjä hakee erityisen relevantteja dokumentteja. Tuloksilla on myös käytännön merkitystä, kun mittarina käytetään Sparck-Jonesin (1974) käytännön merkittävyyttä mittaavaa testiä. Seuraavaksi tutkimuksen tuloksia eritellään tarkemmin kyselyn onnistuneisuuden, relevanssikriteerien ja relevanssipalautteen määrän mukaan.

#### 1.1 Miten aloituskyselyn onnistuneisuus vaikuttaa relevanssipalautteen tuloksellisuuteen

Relevanssipalautteella voidaan parantaa hakutehokkuutta kyselyn onnistuneisuudesta riippumatta. Eritasoisesti onnistuneiden kyselyiden keskinäinen vertailu on hankalaa, johtuen eri vertailukyselyistä, eikä esimerkiksi tilastotestejä tältä ole tältä osin tehty. Näyttää kuitenkin siltä että hyvin onnistuneen aloituskyselyn tuloksellisuus paranee relevanssipalautteen avulla prosenttiyksikköinä mitattuna enemmän kuin heikkotasaisen kyselyn. Heikkotasaisista kyselyä voidaan parantaa parhaimmillaan 12 prosenttiyksikköä verrattuna laajentamattomaan kyselyyn, kun tuloksellisuus lasketaan tiukoilla relevanssikriteereillä. Hyvätasaisen kyselyn tapauksessa päästään jopa 23 prosenttiyksikön tuloksellisuuden lisäykseen.

Relevanssipalautteen toimivuutta eritasoisesti onnistuneilla aloituskyselyillä on tutkittu aika vähän. Käytännössä kyse on kuitenkin osittain siitä, kuinka monta relevanttia dokumenttia palautteeseen pitäisi valita optimaalisen hakutuloksen saavuttamiseksi, mikä puolestaan on huomattavasti tutkitumpi kysymyksen asettelu. Esimerkiksi Keskustalon, Järvelinin ja Pirkolan (2006) tutkimuksessa kävi ilmi, että relevanssipalautteeseen valittavien dokumenttien määrän kasvattaminen parantaa hakutehokkuutta. Tässä tutkimuksessa puolestaan relevanssipalautteella hyvin onnistuneesta aloituskyselystä tuotti enemmän relevantteja dokumentteja kuin heikkotasaisesta kyselystä. Tulos vastaa Keskustalon, Järvelinin ja Pirkolan (2006) tutkimuksen tulosta, sillä hyvin onnistuneen aloituskyselyn tulosjoukossa on enemmän relevantteja dokumentteja kuin heikosti onnistuneen kyselyn, jolloin myös palautteeseen voidaan valita enemmän dokumentteja.

Suhteellisesti laskettuna heikkotasoinen kysely hyötyy kuitenkin eniten relevanssipalautteesta. Esimerkiksi taulukossa 3 esitetty heikosti onnistuneen aloituskyselyn keskitarkkuus on 3,5 prosenttia ja laajennetun kyselyn parhaimmillaan yli 13 prosenttia, jolloin hakutehokkuus kolminkertaistuu aloituskyselyyn nähden. Hyvin onnistuneen laajentamattoman kyselyn keskitarkkuus voi puolestaan olla jopa 27 prosenttia (taulukko 5), mikä tekee hakutehokkuuden kaksinkertaistamisesta melko haastavaa. Tässä mielessä tulos tukee Saltonin ja Buckleyn (1990) tutkimuksen pohjalta tehtyä hypoteesia relevanssipalautteen suotuisasta vaikutuksesta heikosti onnistuneeseen kyselyyn. Kun lähtötaso on heikko, parantamisen varaakin on runsaasti. Lopulta kysymys on kuitenkin siitä, käytetäänkö vertailuissa suhteellisia vai absoluuttisia mittareita.

1.2 Kuinka monta dokumenttia relevanssipalautteeseen pitäisi valita parhaan hakutuloksen saavuttamiseksi?

Relevanssipalautteeseen kannattaa tämän tutkimuksen perusteella valita mahdollisimman monta relevanttia dokumenttia, vaikka hyviin tuloksiin päästään jo yhdelläkin palautedokumentilla. Hyvin ja keskitasoisesti onnistuneilla aloituskyselyillä palautedokumenttien enimmäismäärän kasvattaminen yhdestä viiteen dokumenttiin parantaa useimmiten hakutehokkuutta. Lisäksi tuloksellisuus kohenee monissa tapauksissa käytännön merkittävyyttä mittaavan säännön edellyttämät viisi prosenttiyksikköä. Myös Keskustalo, Järvelin ja Pirkolan (2006) mukaan palautedokumenttien määrän kasvattaminen parantaa hakutehokkuutta. Heidän tutkimuksissaan käyttäjä, joka malttoi selata 30 dokumenttia tuloslistalta ja tehdä relevanssiarvion kaikista dokumenteista, sai myös parhaan hyödyn relevanssipalautteesta (Keskustalo, Järvelin ja Pirkola 2006).

Heikkotasoisilla kyselyillä palautedokumenttien määrällä ei ole merkitystä relevanssipalautteen toimivuuteen. Tämä johtuu siitä, että relevantteja dokumentteja on vain vähän heikkotasoisien kyselyiden tulosjoukossa, jolloin palautedokumenttien määrään ei voida juurikaan kasvattaa.

1.3 Minkä tasoisesti relevantteja dokumentteja relevanssipalautteeseen kannattaa valita, jotta palautteesta saadaan maksimaalinen hyöty?

Käyttäjän kannattaa valita kaikki mahdolliset relevantit dokumentit palautteeseen. Tässä tutkimuksessa palautteeseen käytettyjen dokumenttien relevanssitasolla ei ollut vaikutusta hakutehokkuuteen. Tilastollisesti merkitseviin eroihin olisi mahdollisesti päästy, mikäli

palautedokumenttien määrää olisi kasvatettu suuremmaksi esimerkiksi kasvattamalla alkuperäisen kyselyn tulosjoukkoa kymmenestä 20 tai 30 dokumenttiin, kuten Keskustalo, Järvelin ja Pirkola (2006) olivat tehneet. Suuresta palautedokumenttimäärästä on varaa karsia vähemmän relevantit dokumentit pois. Kun palautteeseen valitaan esimerkiksi vain yksi dokumentti, palautekriteerien välille ei voi tulla eroja, sillä oletuksena on, että myös marginaalisesti relevanteista dokumenteista kiinnostunut käyttäjä valitsee palautteeseen aina relevanteimman dokumentin.

## 2. Pseudorelevanssipalautteen tuloksellisuus

### 2.1 Miten aloituskyselyn onnistuneisuus vaikuttaa pseudorelevanssipalautteen tuloksellisuuteen?

Tämän tutkimuksen perusteella pseudorelevanssipalaute ei yleensä paranna tai heikennä hakutehokkuutta niin, että ero olisi tilastollisesti merkitsevä. Toisin sanoen vaikutus näyttää olevan hyvin sattumanvarainen. Tästä kertoo esimerkiksi se, että taulukossa 5 keskitasoisesti onnistunutta aloituskyselyä onnistuttiin parantamaan tilastollisesti merkitsevästi 3,1 prosenttiyksikköä, kun pseudorelevanssipalautetta annettiin 10 dokumentista. Toisaalta samasta taulukosta nähdään, että hyvin onnistuneeseen kyselyyn pseudorelevanssipalautteella ei ollut vaikutusta, vaikka oletuksena oli, että pseudorelevanssipalaute vaatii toimiakseen onnistuneen aloituskyselyn.

Tutkimustulos ei aivan vastaa esimerkiksi Keskustalon, Järvelinin ja Pirkolan 2006 tutkimusta. He onnistuivat parantamaan hakutehokkuutta pseudorelevanssipalautteella parhaimmillaan lähes neljä prosenttiyksikköä laajentamattomaan kyselyyn nähden. He eivät kuitenkaan laskeneet tulostensa tilastollista merkitsevyyttä, jolloin tulokset saattavat olla osittain sattumaa. (Keskustalo, Järvelin ja Pirkola 2006.) Osittain tulokset kuitenkin tukevat aikaisempaa tutkimusta. Esimerkiksi heikotasoisten kyselyiden tuloksellisuutta pseudorelevanssipalautteen odotetaan heikentävän (esim. Ruthven ja Lalmas 2003; Buckley, Salton, Allan ja Singhal 1995) ja näin kävi myös tässä tutkimuksessa.

### 2.2 Riippuuko pseudorelevanssipalautteen toimivuus palautteeseen valittujen dokumenttien lukumäärästä?

Palautedokumenttien määrällä ei näyttäisi olevan juurikaan merkitystä pseudorelevanssipalautteen toimivuuteen. Erot palautedokumenttien määrän suhteen ovat satunnaisia. Joskin, yhdestä dokumentista annettu pseudopalaute jopa heikensi hakutehokkuutta. Tältä osin tulos tukee osittain

Lundquistin, Grossmanin ja Friederin (1997) tutkimuksen pohjalta tehtyä hypoteesia, jonka mukaan alle viidestä dokumentista annettu palaute toimii heikommin kuin 5-20 dokumentista annettu palaute.

Muissa tutkimuksissa pseudorelevanssipalautetta on testattu niin, että tulosjoukon kärjestä on otettu jopa 30 ensimmäistä dokumenttia mukaan palautteeseen (ks. esim. Buckley et al. 1995 ja Keskustalo, Järvelin ja Pirkola 2006). Ehkä tässäkin tutkimuksessa olisi voitu saada parempia tuloksia pseudorelevanssipalautteen osalta, mikäli palautedokumenttien määrää olisi kasvatettu kymmenestä dokumentista.

## 10. Keskustelua

Tutkimus osoitti, että RATF:n avulla toteutettu relevanssipalaute on tehokas kyselynlaajennusmenetelmä, jonka avulla hakutehokkuutta voidaan parantaa merkittävästi kyselyn onnistuneisuudesta riippumatta. Haun tuloksellisuutta voidaan parantaa helposti yli kymmenen ja parhaimmillaan yli 20 prosenttiyksikköä. Myös aikaisemmissa tutkimuksissa relevanssipalautteen on todettu vaikuttavan positiivisesti hakutehokkuuteen. Miksi relevanssipalautetta ei kuitenkaan ole juuri sovellettu tiedonhakujärjestelmissä?

Relevanssipalaute on kyselynlaajennusmenetelmänä käyttäjän kannalta melko vaivaton ja yksinkertainen. Ainoaksi haasteeksi käyttäjälle jää relevanssiarvioiden tekeminen dokumenteista. Kuitenkaan relevanssipalaute ei ole kovin laajalti käytössä tiedonhakujärjestelmissä. Erityisesti internetin hakukoneissa relevanssipalaute loistaa poissaolollaan. Relevanssipalautteella voidaan parantaa hakutehokkuutta huomattavasti, mutta ehkä relevanssipalautteen aikaansaama hakutehokkuuden lisäys ei kuitenkaan ole tarpeeksi suuri, jotta siitä olisi tarpeeksi hyötyä käyttäjälle. Ensinnäkin relevanssipalautteen avulla voidaan kasvattaa erityisesti saantia (ks. myös Manning, Raghavan ja Schütze 2008, 166), mikä ei ehkä web-hauissa ole kovin tärkeää käyttäjän kannalta. Tulosjoukko paisuu helposti satoihintuhansiin dokumentteihin, jolloin tärkeää olisi saada relevanteimmat dokumentit aivan tuloslistan kärkeen. Toinen syy relevanssipalautteen vähäiselle käyttöönotolle saattaa olla se, että mikäli käyttäjä tuntee hakuaiheen kohtuullisen hyvin jo entuudestaan, myös hakuaiheen muotoilu kyselyksi on helpompaa, eikä relevanssipalautteen apua enää ehkä tarvita. Kolmanneksi relevanssipalautteenantomahdollisuus pitäisi pystyä integroimaan käyttöliittymään siten, että käyttäjän olisi mahdollisimman helppo tarttua siihen. Nopeisiin web-hakuihin tottuneelle käyttäjälle pienikin lisäponnistus voi olla liikaa. Yhtenä relevanssipalautteen sovelluksena voidaan kuitenkin pitää joidenkin internetin hakukoneiden ”Samankaltaisia sivuja” painiketta, joka etsii käyttäjän valitseman sivun kanssa samankaltaisia sivuja. Painikkeen käytöstä tai sen tarkemmasta toimintamekanismista minulla ei kuitenkaan ole tietoa.

Sen sijaan pseudorelevanssipalautteen vaikutus hakutehokkuuteen on hyvin sattumanvarainen. Pseudorelevanssipalautteen hyödyntäminen tiedonhakujärjestelmissä on melko riskialtista puuhaa. Mikäli käyttäjän tekemän kyselyn tuloksellisuutta ei pystytä ennakoimaan tarpeeksi hyvin ja estää heikkoihin kyselyihin tehty palaute, ei pseudorelevanssipalautetta kannata integroida tiedonhakujärjestelmiin. Tämän tutkimuksen valossa näyttäisi siltä, että onnistuneenkaan kyselyn

tuloksellisuutta ei voida parantaa olennaisesti pseudorelevanssipalautteella. Pseudorelevanssipalaute vaatii toimiakseen mahdollisesti vielä enemmän relevantteja dokumentteja tulosjoukkoon.

Tarpeen voisi olla mekanismi, joka pystyisi kohtuullisen luotettavasti ennakoimaan käyttäjän tekemän kyselyn tuloksellisuutta. Näin laajennus voitaisiin jättää tekemättä heikoimpiin kyselyihin. Yksi kyselyn onnistuneisuutta kuvaava indikaattori voisi olla kyselyn pituus. Parin sanan mittaisen kyselyn onnistuneisuutta on hankala ennakoida ja epäonnistumisen riski on suuri. Tällöin myös pseudorelevanssilajennus voi olla viisainta jättää tekemättä. Mikäli pseudorelevanssipalaute jäisi tekemättä, voitaisiin käyttäjää ohjata relevanssipalautteen pariin, sillä jo yhdelläkin palautedokumentilla voidaan parantaa hakutehokkuutta. Pseudorelevanssipalautteen toimivuutta voitaisiin mahdollisesti lisätä myös kasvattamalla palautedokumenttien määrää, kuten jo aiemmin ehdotin. Myös termienvalintamenetelmiin pitäisi ehkä kiinnittää enemmän huomiota. Ehkä tässä tutkimuksessa käytetty RATF ei sittenkään pysty erottelemaan hyviä laajennustermejä huonoista riittävän hyvin. Ainakin kattavampi selvitys tästä voisi olla paikallaan.

Pseudorelevanssipalautteeseen liittyy myös kysymys siitä, tarvitaanko laajennuksen tuomaa (mahdollista) hakutehokkuuden lisäystä, mikäli laajentamaton kyselykin toimii hyvin? Varsinkin, kun odotettavissa oleva hakutehokkuuden lisäys jäänee muutamiin prosenttiyksiköihin. Mikäli käyttäjää kiinnostaa enimmäkseen tuloslistan kärkeen sijoittuneet dokumentit, on näin pieni hakutehokkuuden kasvu merkityksetön.

Tässä tutkimuksessa eritasoiset kyselyt muodostettiin kokeellisesti vastaamaan tutkimuksen tarpeita. Asetelmasta saataisiin realistisempi ja käyttämällä käyttäjien tekemiä kyselyitä. Tämä voitaisiin toteuttaa siten, että joukko käyttäjiä muodostaa kyselyitä hakuaiheista. Sitten kyselyt jaotellaan ryhmiin onnistuneisuuden mukaan. Tämä tekisi koeasetelmasta validimman.

## 11. Lähteet

Baeza-Yates, R. ja Ribeiro-Neto, B. 1999. Modern Information Retrieval. Wokingham: Addison-Wesley.

Buckley, C., Salton, G ja Allan, J. 1994. The effect of adding relevance information in a relevance feedback environment. Teoksessa W. Croft and C. van Rijsbergen (toim.) Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. London: Springer-Verlag, 292-300. [Saatavilla: <http://citeseer.ist.psu.edu/buckley94effect.html> Viitattu 4.6.2008.]

Buckley, C., Salton, G., Allan, J. ja Singhal, A. 1995. Automatic query expansion using SMART: TREC-3. Teoksessa D. K. Harman (toim.) Proceedings of the Third Text Retrieval Conference (TREC-3). NIST special publication 500-225, 69-80

Callan, J. P., Croft, W. B. ja Harding, S. M. 1992. The INQUERY retrieval system. Teoksessa Proceedings of the Third International Conference on Database and Expert Systems Applications. Valencia: Springer-Verlag, 78-83

Chang, Y. K., Cirillo, C. ja Razon, J. 1971. Evaluation of feedback retrieval using modified freezing, residual collection and test and control groups. Teoksessa Gerald Salton (toim.) The SMART retrieval system - experiments in automatic document processing. Montvale: Prentice-Hall, 355-370.

Conover, W. J. 1980. Practical nonparametric statistics. 2. p. New York: John Wiley & Sons.

Efthimiadis, E. N. 1995. User Choices: A New Yardstick for the Evaluation of Ranking Algorithms for Interactive Query Expansion. Information Processing & Management, 31 (4), 605-620.

Efthimiadis, E. N. 1996. Query expansion. Teoksessa M. E. Williams (toim.) Annual review of information science and technology (ARIST) 31. Medford: Learned Information for the American



Society for Information Science, 121-187. [Saatavilla:

<http://faculty.washington.edu/efthimis/pubs/Pubs/qe-arist/QE-arist.html> Viitattu 4.6.2008]

Järvelin, K. 1995. Tekstiedonhaku tietokannoista : Johdatus periaatteisiin ja menetelmiin. Espoo: Suomen ATK-kustannus.

Kekäläinen, J. 1999. The effects of query complexity, expansion and structure on retrieval performance in probabilistic text retrieval. Väitöskirja, informaatiotutkimuksen laitos, Tampereen yliopisto. Acta Universitatis Tamperensis 678. Tampere: Tampereen yliopisto. [Saatavilla: <http://www.info.uta.fi/tutkimus/fire/archive/QCES.pdf>. Viitattu 15.4.2008.]

Kekäläinen, J. ja Järvelin, K. 2002. Using graded relevance assessments in IR evaluation. Journal of the American Society for Information Science and Technology 53 (13), 1120-1129. [Saatavilla: <http://www.info.uta.fi/tutkimus/fire/archive/JKKJ-gPR.pdf> Viitattu 6.5.2008.]

Keskustalo, H., Järvelin, K. ja Pirkola, A. 2006. The effects of relevance feedback quality and quantity in interactive relevance feedback: a simulation based on user modelling. Teoksessa Lecture Notes in Computer Science vol. 3936. Berlin: Springer 191-204. [Saatavilla: [www.info.uta.fi/tutkimus/fire/archive/2006/Keskustalo-ECIR06.pdf](http://www.info.uta.fi/tutkimus/fire/archive/2006/Keskustalo-ECIR06.pdf). Viitattu 21.4.2008.]

Kwok, K. L. 1996. A new method of weighting query terms for ad-hoc retrieval. Teoksessa Proceedings of the 19th annual international ACM SIGIR conference on research and development in information retrieval. New York: ACM, 187-195.

Lin, J. 2006. Relevance feedback, luentomateriaali. College of Information Studies. University of Maryland. [Saatavilla: <http://www.umiacs.umd.edu/~jimmylin/LBSC796-INFM718R-2006-Spring/lecture7.ppt>. Viitattu 6.5.2008.]

Lundquist, C., Grossman, D. A. ja Frieder, O. 1997. Improving relevance feedback in the vector space model. Teoksessa CIKM '97: proceedings of the sixth international conference on Information and knowledge management. New York: ACM, 16-23.

Magennis, M. ja van Rijsbergen, C. J. 1997. The potential and actual effectiveness of interactive query expansion. SIGIR Forum 31, 324-332.

Manning, C. D., Raghavan, P. ja Schütze, H. 2008. Introduction to information retrieval. Cambridge University Press: New York. [Saatavilla <http://www-csli.stanford.edu/~hinrich/information-retrieval-book.html> Viitattu 7.9.2008.]

Metzler, D., Strohman, T., Turtle, H. ja Croft, W. B. 2004. Indri at trec 2004: Terabyte track. Teoksessa Proceedings TREC 2004

Metzler, D. 2005. Indri retrieval model overview.  
<http://ciir.cs.umass.edu/~metzler/indriretmodel.html> (käytetty 21.9.2008).

Pirkola, A., Leppänen, E., ja Järvelin, K. 2002. The RATF formula (Kwok's formula): exploiting average term frequency in cross-language retrieval. Information Research 7 (2). [Saatavilla <http://informationr.net/ir/7-2/infres72.html>. Viitattu 15.4.2008.]

Robertson, S. E. 1990. On term selection for query expansion. Journal of Documentation. 46 (4), 359-364.

Rocchio, J. J. 1971. Relevance feedback in information retrieval. The SMART retrieval system - experiments in automatic document processing. Teoksessa Gerald. Salton (toim.) The SMART retrieval system - experiments in automatic document processing. Montvale: Prentice-Hall, 313-323.

Ruthven, I., Lalmas, M. 2003. A survey on the use of relevance feedback for information access systems. Knowledge Engineering Review 18 (2), 95-145. [Saatavilla [http://inex.is.informatik.uni-duisburg.de:2004/pdf/ker\\_ruthven\\_lalmas.pdf](http://inex.is.informatik.uni-duisburg.de:2004/pdf/ker_ruthven_lalmas.pdf) Viitattu 6.5.2008.]

Salton, G. ja Buckley, C. 1988. Term weighting approaches in automatic text retrieval. Information Processing & Management 24 (5), 513-523.

Salton, G. ja Buckley, C. 1990. Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science 41(4), 288-297. [Saatavilla: <http://www.umiaccs.umd.edu/~jimmylin/LBSC796-INF718R-2006-Spring/papers/Salton90.pdf>. Viitattu 2.7.2008.]

Siegel, S. ja Castellan N. J. 1988. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.

Sormunen, E. 2002a. Liberal Relevance Criteria of TREC - Counting on Negligible Documents? Teoksessa Beaulieu, M. et al. (toim.) Proceedings of the Twenty-Fifth Annual ACM SIGIR Conference on Research and Development in Information Retrieval. August 11-15, 2002, Tampere, Finland. Special Issue of SIGIR Forum 36():324-330. [Saatavilla: <http://www.info.uta.fi/tutkimus/fire/archive/298-sormunen.pdf> Viitattu 23.5.2008.]

Sormunen, E. 2002b. A Retrospective Evaluation Method for Exact-Match and Best-Match Queries Applying an Interactive Query Performance Analyser. 24th BCS-IRSG European Colloquium on IR Research March 25-27, 2002, Glasgow, Scotland, UK.

Sparck Jones, Karen. 1974. Automatic indexing. Journal of documentation 30 (4), 393-432.

Strohman, T., Metzler, D., Turtle, H., and Croft, W. B. 2005. Indri: A language-model based search engine for complex queries (extended version)" CIIR Technical Report. [Saatavilla: <http://ciir.cs.umass.edu/pubfiles/ir-407.pdf> Viitattu 23.5.2008.]

## **Liite**

### ***Merkitsevyydestien tulokset***

Taulukoissa 1-9 on esitetty laajennusten välisten merkitsevyydestien tulokset. Taulukoissa on huomioitu ainoastaan sellaiset laajennukset, jotka eroavat tilastollisesti merkitsevästi ainakin yhdestä toisesta laajennuksesta. \*\*\* tarkoittaa, että laajennusten välinen ero on tilastollisesti erittäin merkitsevä ja \*\* viittaa tilastolliseen merkitsevyyteen. (Ks. testeistä lisää luvusta 8.5.2.) Esimerkiksi taulukossa 1 on esitetty ainoastaan relevanssipalautemenetelmien vertailu pseudorelevanssipalautemenetelmiin, sillä relevanssipalautemenetelmien välille ei saatu tilastollisesti merkitseviä eroja. Sen sijaan esimerkiksi taulukossa 2 relevanssipalautetta yhdestä dokumentista on verrattu muihin relevanssipalautemenetelmiin, sillä se osoittautui tilastollisesti heikommaksi kuin muut menetelmät. Tämän vuoksi taulukot ovat erikokoisia, eikä kaikkien menetelmien välisiä tuloksia esitetä.

Taulukoissa sarakkeen 1 laajennusta on verrattu rivin 1 laajennuksiin. Esimerkiksi taulukossa 1 yhdestä dokumentista liberaaleilla kriteereillä annettu palaute on tilastollisesti erittäin merkitsevästi parempi kuin yhdestä, viidestä tai kymmenestä dokumentista annettu pseudorelevanssipalaute. Ero pseudorelevanssipalautteeseen on 9,3; 8,0 ja 7,7 prosenttiyksikköä tässä järjestyksessä. Itse asiassa kaikki taulukon 1 relevanssipalautelaajennukset ovat tilastollisesti erittäin merkitsevästi parempia kuin yksinään pseudorelevanssilajennus. Miinusmerkki lukeman edessä tarkoittaa, että rivillä esitetty laajennus on parempi kuin sarakkeen laajennus. Esimerkiksi taulukossa 1 yhdestä dokumentista tehty pseudorelevanssilajennus on tilastollisesti merkitsevästi huonompi kuin viidestä dokumentista tehty laajennus. Ero menetelmien välillä on 1,3 prosenttiyksikköä. Vertailut laajentamattomaan kyselyyn on esitetty luvun 9 taulukoissa. Liitteen taulukot on järjestelty samaan tapaan kuin luvussa 9. Ensinnäkin järjestysperusteena on se, millä relevanssikriteereillä tuloksellisuus on laskettu ja toiseksi se, miten kysely on onnistunut.

**Laajennusmenetelmien vertailu, tuloksellisuus laskettu liberaaleilla relevanssikriteereillä.**

**Taulukko 1. Heikosti onnistuneen kyselyn laajennusten vertailu**

| Heikko | 1PRF   | 5PRF   | 10PRF  |
|--------|--------|--------|--------|
| 1RF1   | 9,3*** | 8,0*** | 7,7*** |
| 5RF1   | 10***  | 8,7*** | 8,4*** |
| 1RF2   | 8,4*** | 7,0*** | 6,8*** |
| 5RF2   | 8,7*** | 7,3*** | 7,0*** |
| 1RF3   | 7,7*** | 6,4*** | 6,1*** |
| 1PRF   |        | -1,3** |        |

**Taulukko 2. Keskitasoisesti onnistuneen kyselyn laajennusten vertailu**

| Keski123 | 5RF1    | 10RF1   | 5RF2    | 1RF3   | 5RF3   | 1PRF    | 5PRF    | 10PRF   |
|----------|---------|---------|---------|--------|--------|---------|---------|---------|
| 1RF1     | -5,1*** | -5,2*** | -4,0*** |        |        | 9,9***  | 5,9***  | 6,0***  |
| 5RF1     |         |         |         | 5,3*** | 3,5*** | 15,2*** | 11,2*** | 11,3*** |
| 10RF1    |         |         |         | 5,5*** | 3,7*** | 15,4*** | 11,4*** | 11,5*** |
| 5RF2     |         |         |         | 4,3*** | 2,4**  | 14,1*** | 10,2*** | 10,2*** |
| 1RF3     |         |         |         |        |        | 9,9***  | 5,9**   | 6,0**   |
| 5RF3     |         |         |         |        |        | 11,7*** | 7,7***  | 7,8***  |
| 1PRF     |         |         |         |        |        |         | -4,0**  | -3,9**  |

**Taulukko 3. Hyvin onnistuneen kyselyn laajennusten vertailu**

| Hyvä123 | 5RF1    | 10RF1   | 5RF2    | 10RF2   | 5RF3   | 1PRF    | 5PRF    | 10PRF   |
|---------|---------|---------|---------|---------|--------|---------|---------|---------|
| 1RF1    | -5,2*** | -5,7*** | -5,0*** | -5,1*** | -2,8** | 11,2*** | 6,0***  | 8,8***  |
| 5RF1    |         |         |         |         |        | 16,2*** | 11,2*** | 14,0*** |
| 10RF1   |         |         |         |         |        | 16,9*** | 11,7*** | 14,5*** |
| 5RF2    |         |         |         |         |        | 16,1*** | 10,9*** | 13,8*** |
| 10RF2   |         |         |         |         |        | 16,3*** | 11,1*** | 14,0*** |
| 5RF3    |         |         |         |         |        | 14,0*** | 8,8***  | 11,6*** |
| 1PRF    |         |         |         |         |        |         | -5,2**  |         |

## Laajennusmenetelmien vertailu, tuloksellisuus laskettu keskitiukoilla relevanssikriteereillä

**Taulukko 4. Heikosti onnistuneen kyselyn laajennusten vertailu**

| Heikko23 | 1PRF    | 5PRF   | 10PRF  |
|----------|---------|--------|--------|
| 1RF1     | 10,3*** | 9,5*** | 8,9*** |
| 5RF1     | 9,5***  | 8,7*** | 8,1*** |
| 1RF2     | 9,6***  | 8,8*** | 8,3*** |
| 5RF2     | 10,0*** | 9,2*** | 8,6*** |
| 1RF3     | 8,9***  | 8,1*** | 7,5*** |
| 1PRF     |         |        | -1,4** |

**Taulukko 5. Keskitasoisesti onnistuneen kyselyn laajennusten vertailu**

| Keski23 | 5RF1    | 10RF1   | 5RF2    | 1RF3   | 5RF3  | 1PRF     | 5PRF    | 10PRF   |
|---------|---------|---------|---------|--------|-------|----------|---------|---------|
| 1RF1    | -4,9*** | -5,1*** | -4,7*** |        |       | 10,8***  | 7,3**   | 6**     |
| 5RF1    |         |         |         | 5,0*** | 2,9** | 15,7***  | 12,2*** | 10,9*** |
| 10RF1   |         |         |         | 5,2*** | 3,0** | 15,8***  | 12,3*** | 11,1*** |
| 5RF2    |         |         |         | 4,8*** | 2,7** | 15,5***  | 12,0*** | 10,7*** |
| 1RF3    |         |         |         |        |       | 10,67*** | 7,2**   |         |
| 5RF3    |         |         |         |        |       | 12,8***  | 9,4***  | 8,1***  |
| 1PRF    |         |         |         |        |       |          | -3,5*** | -4,8*** |

**Taulukko 6. Hyvin onnistuneen kyselyn laajennusten vertailu**

| Hyvä23 | 5RF1    | 10RF1   | 5RF2    | 10RF2   | 5RF3   | 1PRF    | 5PRF    | 10PRF   |
|--------|---------|---------|---------|---------|--------|---------|---------|---------|
| 1RF1   | -4,9*** | -4,8*** | -5,6*** | -5,8*** | -2,9** | 15,8*** | 10,5*** | 12,6*** |
| 5RF1   |         |         |         |         |        | 20,3*** | 15,4*** | 17,5*** |
| 10RF1  |         |         |         |         |        | 20,6*** | 15,3*** | 17,4*** |
| 5RF2   |         |         |         |         |        | 21,4*** | 16,1*** | 18,2*** |
| 10RF2  |         |         |         |         |        | 21,6*** | 16,3*** | 18,4*** |
| 5RF3   |         |         |         |         |        | 18,8*** | 13,4*** | 15,5*** |

## Laajennusmenetelmien vertailu, tuloksellisuus laskettu tiukoilla relevanssikriteereillä

**Taulukko 7. Heikosti onnistuneen kyselyn laajennusten vertailu**

| Heikko3 | 1PRF    | 5PRF    | 10PRF   |
|---------|---------|---------|---------|
| 1RF1    | 14,5*** | 13,4*** | 13,7*** |
| 5RF1    | 12,1*** | 11,0*** | 11,3*** |
| 1RF2    | 14,4*** | 13,3*** | 13,6*** |
| 5RF2    | 14,6*** | 13,5*** | 13,8*** |
| 1RF3    | 13,8*** | 12,7*** | 13,0*** |

**Taulukko 8. Keskitasoisesti onnistuneen kyselyn laajennusten vertailu**

| Keski3 | 5RF3   | 1PRF    | 5PRF    | 10PRF   |
|--------|--------|---------|---------|---------|
| 1RF1   | -3,0** | 12,8*** | 8,18**  |         |
| 5RF1   |        | 15,1*** | 10,5*** | 10,3*** |
| 10RF1  |        | 15,2*** | 10,6*** | 10,4*** |
| 5RF2   |        | 15,1*** | 10,5*** | 10,2*** |
| 1RF3   |        | 12,9*** | 8,3***  | 8,1**   |
| 5RF3   |        | 15,7*** | 11,1*** | 10,9*** |
| 1PRF   |        |         | -4,6*** | -4,9*** |

**Taulukko 9. Hyvin onnistuneen kyselyn laajennusten vertailu**

| Hyvä3 | 5RF3   | 1PRF    | 5PRF    | 10PRF   |
|-------|--------|---------|---------|---------|
| 1RF1  | -3,8** | 21,7*** | 15,4*** | 20,9*** |
| 5RF1  |        | 22,5*** | 16,2*** | 21,7*** |
| 10RF1 |        | 18,6*** | 12,3*** | 17,8*** |
| 5RF2  |        | 19,3*** | 13***   | 18,5*** |
| 10RF2 |        | 19,3*** | 13,0*** | 18,5*** |
| 5RF3  |        | 25,5*** | 19,2*** | 24,7*** |