

Eero Sormunen

# Retrospektiivinen menetelmä Boolean kyselyjen tehokkuuden mittaamiseen\*

Eero Sormunen, Retrospektiivinen menetelmä Boolean kyselyjen tehokkuuden mittaamiseen. [A Retrospective Method for Measuring the Effectiveness of Boolean Queries] Informaatiotutkimus 19 (3): 74-81, 2000.

Traditional methods for the system-oriented evaluation of Boolean IR systems suffer from validity and reliability problems. Laboratory-based research neglects the searcher and studies suboptimal queries. Research on operational systems fails to make a distinction between searcher performance and system performance. This approach is neither capable of measuring performance at standard points of operation (e.g. across recall levels  $R_{0,0}$ - $R_{1,0}$ ).

A new laboratory-based evaluation method for Boolean IR systems is proposed. It is based on a controlled formulation of inclusive query plans, on an automatic conversion of query plans into elementary queries, and on combining elementary queries into optimal queries at standard points of operation. The method is retrospective since full relevance data are applied in the optimization of queries. Major results of a large case experiment are reported.

*Address: Eero Sormunen, University of Tampere, Departments of Information Studies, FIN -33014 UNIVERSITY OF TAMPERE, Finland. Email: eero.sormunen@uta.fi*

## 1 Johdanto

Boolean malliin perustuvat tiedonhakujärjestelmät ovat olleet pääroolissa käytännön sovelluksissa lähes 40 vuoden ajan. Vasta Internetin hakupalvelujen kehityksen myötä osittaistäsmäytykseen perustuvat hakujärjestelmät ovat saaneet näkyvämmän aseman. Boolean mallin sovellusten yleisyydestä huolimatta niitä on tutkittu hämmästyttävän vähän eikä Boolean mallin ominaisuuksia kyselyiden ja dokumenttien täsmäyttäjänä tunneta kovin hyvin. Tilanteeseen on nähtävissä useita selittäviä tekijöitä:

1) Perinteisillä tutkimusmenetelmällä ei ole pystytty selkeästi erottamaan hakijan ja teknisen järjestelmän osuutta havainnoiduissa ilmiöissä. Boolean kyselyjen muotoilu edellyttää asiantuntevaa hakijaa, jotta hakuetojen määrittely tapahtuu mielekkäästi (Ingwersen & Willet 1995). Kun hakijaa ei ole osattu sijoittaa koeasetelmaan kontrolloidusti, hänet on joko jätetty kokonaan pois tutkittavasta hakuprosessista (laboratoriomallin tutkimus) tai hakijan vaikutus on jätetty kontrolloimatta (operatiionaalisten järjestelmien tutkimus). Molemmat vaihtoehdot ovat johtaneet Boolean mallin osalta epäkelvoinen evaluointituloksiin.

2) Boolean kysely poimii tietokannasta ne dokumentit, jotka täsmäävät tarkalleen kyselyn ehtoihin. Käyttäjien tarpeet kuitenkin vaihtelevat eri tilanteissa. Joskus halutaan löytää kaikki aihepiiriin dokumentit ja ollaan valmiita selaamaan suuriakin tulosjoukkoja. Joskus hakijalle riittää

---

\*Artikkeli perustuu Sormusen väitöskirjaan (Sormunen 2000).

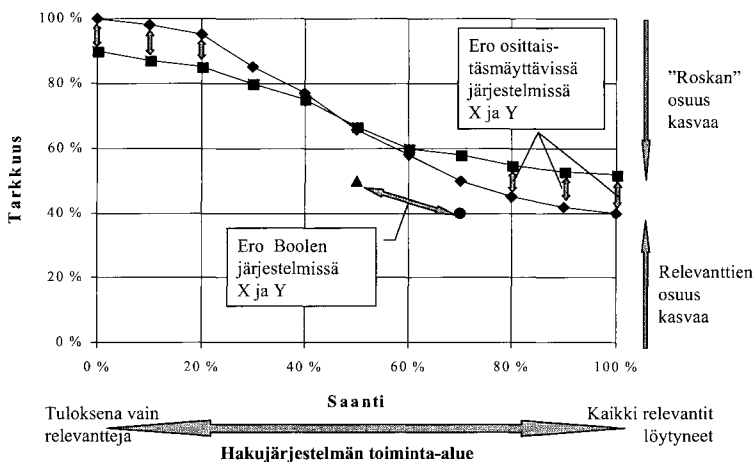
muutama relevantti dokumentti ja epärelevantit dokumentit halutaan torjua kokonaan tulosjoukosta. Eri tavoitteita vastaava tulos edellyttää Boolean järjestelmässä aina uuden kyselyn muotoilua. Osittaismäyttävissä järjestelmissä eri tavoitteita vastaava tulos löydetään selaamalla riittävän pitkälle yhden kyselyn relevanssilajiteltua tuloslistaa. Tästä Boolean järjestelmän erityispiirteestä seuraa hyvin hankala tehokkuuden mittaongelma, jota tutkijat eivät ole pystyneet tyydyttävästi tätä ennen ratkaisemaan. Ongelmaan palataan tarkemmin alempana.

3) Järjestelmäsuuntautuneen tiedonhaku tutkimuksen päävirta on jo 1960-luvulta keskittynyt osittaismäyttävien hakumenetelmien tutkimukseen, koska Boolean mallia on pidetty väistyvänä teknologiana (Frants ym. 1999). Asenteen seuraukset näkyvät vähäisenä Boolean järjestelmien tutkimuksena myös Boolean järjestelmien tutkimukseen soveltuvien menetelmien kehityksen laiminlyöntinä. Boolean järjestelmien säilyminen vallitsevana teknologiana käytännön sovelluksissa (www-palveluja lukuun ottamatta) ja kohtuullisen hyvä menestyminen vertailutesteissä viittaavat siihen, että väistyvässä teknologiassakin on jotain kilpailukykyisiä piirteitä. Boolean mallin vahvoja piirteitä ei kuitenkaan ole näissä tutkimuksissa pystytty selkeästi tunnistamaan (Paris & Tibbo 1998, Hersh & Hickam 1995).

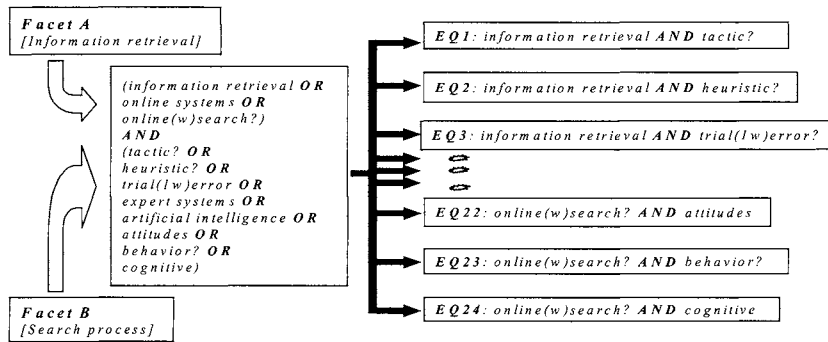
**Kuva 1** havainnollistaa Boolean järjestelmien tehokkuuden mittaamiseen liittyvää ongelmaa. Osittaismäyttävien hakumenetelmien vertailussa voidaan kyselytuloksen tarkkuus laskea jokaisen kyselyn relevanssilajittelusta tuloslistasta hakujärjestelmän toiminta-alueen vakioasteissa,

esimerkiksi 10 %:n välein saantiasteikolla  $R_{0,1} - R_{1,0}$ . Eri testitehtävissä saavutettua tarkkuutta voidaan verrata jossakin vakioasteissa järjestelmän toiminta-alueen eri osissa. Eri hakumenetelmien keskimääräistä tehokkuutta voidaan nyt verrata itsenäisesti toiminta-alueen eri osissa laskemalla tarkkuuden keskiarvo valituilla saantitasolla yli kaikkien testihakujen. Esimerkiksi kuvan 1 osittaismäyttävien hakumenetelmien erot on helppo todeta. Toinen menetelmä on parempi hakujärjestelmän alemmalla (tavoite löytää muutama relevantti dokumentti), toinen hakujärjestelmän ylemmällä toimintakäyrän alueella (tavoitteena löytää kaikki relevantit).

Myös Boolean järjestelmien tutkimuksessa on perinteisesti tehty vain yksi kysely kustakin testitehtävästä ja tuon kyselyn tulosjoukosta on laskettu yksi lukupari: saanti ja tarkkuus. Käytännössä eri testitehtävistä muotoillut kyselyt edustavat hakujärjestelmän toiminta-alueen eri osia. Kun saannin ja tarkkuuden keskiarvot lasketaan erikseen, lopputulos on kaksi keskiarvolukua järjestelmän toiminta-alueen keski-vaiheilla. Keskiarvolukujen pohjalta on usein mahdotonta tehdä järjeviä päätelmiä verrattavien Boolean kyselyjen tehokkuuseroista. Esimerkissä kahden Boolean menetelmän X ja Y keskimääräinen tehokkuus voi hyvin edustaa saman veroisten menetelmien toiminta-alueen eri osia tai pientä eroa menetelmien tehokkuudessa.



**Kuva 1.** Perinteinen tapa verrata tiedonhakumenetelmien keskimääräistä tehokkuutta Boolean ja relevanssilajitelluissa hakujärjestelmissä.



Kuva 2. Kattava hakusuunnitelma ja muunnos alkeiskyselyiksi (EQ = elementary query).

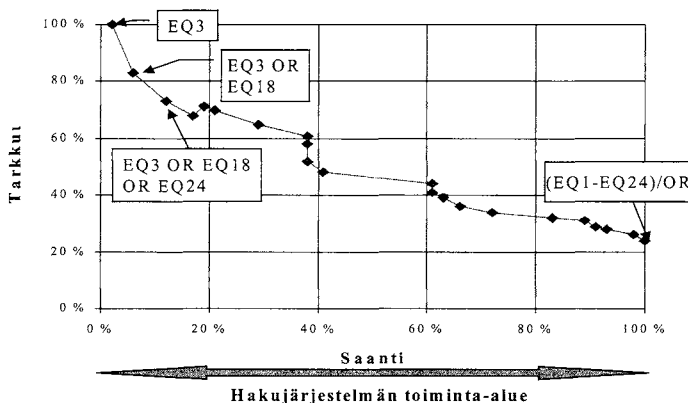
## Harterin idea

Stephen Harter (1990) esitti idean menetelmästä, jolla Boolean kyselyjen tehokkuusvertailuissa voitaisiin välttää edellä kuvattuja mittausongelmia. Hänen lähestymistapansa perustui alkeiskyselyiden (elementary queries) käyttöön. Hän esitteli idean käyttämällä esimerkkinä yhtä hakutehtävää, jonka tavoitteeksi oli määritelty löytää dokumentteja tiedonhakuprosessin eri ulottuvuuksiin liittyvästä tutkimuksesta. Harterin ideaan liittyvän hakutehtävän käsittelyvaiheita on havainnollistettu kuvissa 2 ja 3.

Ensimmäiseksi Harter käytti ammattihakijoille tuttua tapaa laatia kattava hakusuunnitelma ns. lohkostrategiaa käyttäen (kuva 2). Ensin tunnistetaan hakutehtävän pääkäsitteet (fasetit) ja sitten ideoidaan millä kaikilla ilmauksilla dokumentin kirjoittaja on saattanut viitata näihin

käsitteisiin. Sen jälkeen Harter suoritti kyselyn tietokannassa ja arvioi kaikkien 371 löytyneen dokumentin relevanssin suhteessa tehtävän aiheeseen. Seuraavaksi Harter hajotti hakusuunnitelmansa alkeiskyselyiksi, jotka muodostuvat kombinoimalla vuorotellen kahteen fasettiin kuuluneet hakusanat. Näin syntyi  $3 \times 8 = 24$  alkeiskyselyä.

Käyttämällä hyväksi relevanssitietoja voidaan kunkin alkeiskyselyn tehokkuus laskea. Kun tiedetään kuinka monta relevanttia ja epärelevanttia dokumenttia kukin alkeiskysely löytää, on helppo vertailla alkeiskyselyiden keskinäistä paremmuutta. Alkeiskyselyistä voidaan muodostaa yhdistelmiä, jotka johtavat parhaaseen mahdolliseen tarkkuuteen toiminta-alueen eri osissa (eri saantitasoilla).



Kuva 3. Esimerkkihaun optimaalisten kyselyjen tarkkuus saantitasoilla  $R_{0,1}$ - $R_{1,0}$  (Harterin esimerkki).

Harter esitteli yksinkertaisen heuristisen algoritmin, jolla alkeiskyselyjä yhdisteltiin yksi kerrallaan (OR-operaattorilla) disjunktiviseksi kyselyksi. Ensimmäiseksi optimaaliseksi valittiin suurimpaan tarkkuuteen johtanut alkeiskysely. Yhdistelmään lisättiin aina uusi alkeiskysely niin, että uuden yhdistelmän tuottama tarkkuus maksimoitui saavutetulla saantitasolla. Eri yhdistelmäkyselyiden tarkkuusarvot on esitetty saannin funktiona kuvassa 3. Harterin esimerkki havainnollistaa, että Boolean kyselyjäkin voidaan arvioida hakujärjestelmän eri toimintalueilla. Hän esitti kuitenkin idean vain luonnosmaisesti. Ideasta on helppo löytää metodologisia puutteita, eikä sitä esitetyssä muodossa voi soveltaa käytännön tutkimusmenetelmänä.

## 2 Tutkimusongelmat

Väitöskirjatutkimuksen tavoitteena oli kehittää Harterin idean pohjalta evaluointimenetelmä Boolean tiedonhakujärjestelmien toiminnallisen tehokkuuden mittaamiseen järjestelmäsuuntautuneesta näkökulmasta huomioiden kuitenkin asiantuntevan hakijan rooli kyselyjen suunnittelussa. Menetelmän kuvaamisessa ja arvioinnissa sovellettiin Nevellin (1969) viitekehystä, jonka mukaan menetelmän systemaattiseen kuvaukseen kuuluu kolme pääkomponenttia:

1) Menetelmän sovellusalan (domain) määrittely. Tavoitteena on vastata kysymykseen, minkä tyyppiseen tiedonhakujärjestelmien evaluointiin kehitettävä menetelmä soveltuu ja on erityisen hyvä.

2) Menetelmän työvaiheiden (procedure) kuvaus. Menetelmän käyttöön liittyvät operaatiot on kuvattava niin yksiselitteisesti, että eri tutkijat pystyvät soveltamaan sitä yhdenmukaisesti.

3) Menetelmän perustelujen (justification) esittäminen. Menetelmän on mahdollistettava kiinnostavien ja mielekkäiden evaluointitehtävien suorittaminen (appropriateness), sen on täytettävä tieteellisen tutkimuksen validisuus- ja luotettavuuskriteerit sekä menetelmän on oltava tutkimusekonomisesti käyttökelpoinen (ks. Tague-Sutcliffe 1992, Saracevic 1995).

## 3 Menetelmän kuvaus

### Sovellusala

Esitetyn menetelmän keskeinen erikoispiirre on kyselyjen optimointi relevanssitetöiden perusteella. Robertson (1996) kutsuu relevanssitetöitä hyödyntäviä menetelmiä retrospektiivisiksi erotuksena tyypillisistä

evaluointimenetelmistä jotka ovat ennustavia (predictive). Ennustavissa tiedonhaun evaluointimenetelmissä pyritään jäljittelemään käytännön hakutilannetta (relevantit dokumentit eivät etukäteen tiedossa), kun retrospektiivisessä lähestymistavassa hakijan ja järjestelmän toiminta idealisoidaan. Tällöin pyritään vastaamaan kysymykseen, mikä on (teknisen) tiedonhakujärjestelmän tehokkuus, jos sen toimintakapasiteettia voitaisiin hyödyntää maksimaalisesti.

Menetelmä on tarkoitettu Boolean mallin ja siihen perustuvien hakumenetelmien evaluoivaan tutkimukseen ja väitöskirjassa tarkastelu rajattiin vapaatekstihakuun. Tällöin on kiinnostuksen kohteena Boolean kyselyiden ja tekstidokumenttien täsmäyttämisen ilmiöt. Boolean kyselyiden muotoilussa on kyselyn rakenteella hyvin tärkeä rooli ja menetelmä mahdollistaa kyselyn rakenteellisten piirteiden (tyhjentävyys, kattavuus) ja järjestelmän toiminta-alueen eri tasoilla saavutetun hakutehokkuuden vertailun.

Voidaan esimerkiksi selvittää miten optimaaliseen tulokseen johtavat kyselyt poikkeavat rakenteellisesti toisistaan pienissä ja suurissa tietokannoissa, suureen tarkkuuteen ja suureen saantiin tähtäävissä hauissa, eri tyyppisiä dokumentteja haettaessa, erilaisia operaattoreita käytettäessä (esimerkiksi läheisyysoperaattorit vs. AND-operaattori), jne. Ks. Sormunen (2000, 55-58; 195-198).

### Menetelmän työvaiheet

Menetelmän käytännön soveltamisen työvaiheet jakaantuvat kolmeen pääjaksoon: (1) hakusuunnitelmien laadintaan, (2) kyselyjen optimointiin ja (3) tulosten analysointiin.

1) *Hakusuunnitelmien laadinta*. Testihakutehtävät annetaan yhden tai useamman ammattihakijan analysoitavaksi ja he laativat niistä kattavat hakusuunnitelmat. Haun suunnittelijat keskittyvät tunnistamaan kaikki haun kannalta kiinnostavat fasetit (lohkot) ja kaikki fasetteja edustavat hakusanat. Kattavat hakusuunnitelmat kuvaavat periaatteessa kaikki vaihtoehtoiset järjestyvät kyselyt annetusta hakutehtävästä. Kattavien hakusuunnitelmien perusteella voidaan muodostaa ns. kyselyjen säätelyvaruus, jonka puitteissa on mahdollista tutkia miltä tyhjentävyytasolta (fasettien eli rajausten määrä) ja kattavuustasolta (hakusanojen määrä per fasetti) optimaalisesti toimiva kysely löytyy. Perinteisessä evaluoinnissa hakusuunnitelmat ovat perustuneet kunkin ammattihakijan mutu-käsityksiin eikä hakijan vaikutusta tuloksiin ole voitu kontrolloida. Harterin esimerkistä poiketen (kuva 2) hakusuunnitelma laaditaan kaikilla mahdollisilla tyhjentävyytasooilla 1-n, missä n on hakutehtävästä tunnistettujen fasettien määrä.

2) *Kyselyjen optimointi*. Parhaiten toimivien kyselyjen löytäminen kyselyjen säätelyvaruudesta perustuu kahteen automaattiseen prosessiin. Ensin hakusuunnitelmat pilkotaan alkeiskyselyiksi (kuten kuvassa 2). Alkeiskyselyistä koostetaan kehitetyin optimointialgoritmin avulla parhaiten toimiva yhdistelmä kussakin valitussa vakiotoimintapisteessä (standard point of operation). Vakiotoimintapisteinä käytettiin sekä kiinteitä saantitasoja  $R_{0,1}$  -  $R_{1,0}$  että kiinteitä tulosjoukon maksimikokoja (document cut-off value - DCV). Kyselyjen optimointialgoritmi kehitettiin operaatiotutkimuksen erään tyyppitehtävän, ns. binäärisen repuntäytöngelman (0-1 Knapsack Problem) tunnettujen ratkaisumallien pohjalta.

3) *Tulosten analysointi*. Prosessi tuottaa lohkorakenteen kaltaisia kyselyjä, jotka edustavat (estimoivat) optimaalisesti toimivaa kyselyä kyseisessä toimintapisteessä (saantitaso, tulosjoukon koko). Vertaamalla optimaalisten kyselyjen tuottamia tarkkuusarvoja voidaan päätellä, onko eri hakumenetelmien välillä tehokkuuseroja. Vastaavasti voidaan analysoida minkälaisia ovat optimaalisten kyselyiden rakenteet (tyhjentävyys, kattavuus) vertailtavissa hakutilanteissa. (Sormunen 2000, 25-54).

## Käyttökelpoisuuden perustelut

Väitöskirjatutkimuksessa tarkasteltiin perusteellisesti uuden evaluointimenetelmän käyttökelpoisuuden eri aspekteja mm. suorittamalla evaluointitutkimus, tekemällä menetelmän eri työvaiheisiin liittyviä validisuus- ja luotettavuustestejä sekä arvioimalla analyttisin keinoin menetelmän tutkimusekonomista tehokkuutta. Hakusuunnittelun luotettavuutta selvitettiin mm. testaamalla eri henkilöiden hakufasettien valinnan yhdenmukaisuutta. Optimoinnin luotettavuutta testattiin toteuttamalla kilpaileva optimointi tiedonhakupelin avulla, jolloin asiaan perehtynyt tiedonhakijan pyrki pelin avulla löytämään optimointialgoritmia parempia kyselyjä. Tarkastelu osoitti, että menetelmään ei liity merkittäviä validisuus-, luotettavuus- tai tehokkuusongelmia (Sormunen 2000, 153-194).

## 4 Esimerkkitutkimus

Väitöskirjatutkimuksessa on raportoitu laaja esimerkkitutkimus, jossa uutta menetelmää soveltamalla selvitettiin Boolean kyselyiden tehokkuuteen ja rakenteisiin vaikuttavia ilmiöitä suurissa tekstitietokannoissa. Tarkastelussa oli tietokannan koon lisäksi relevanttien dokumenttien yleisyys (generality) eli tiheys (density). Pienessä tietokannassa optimoituja kyselyjä verrattiin kahdessa suuressa tietokannassa optimoituihin kyselyihin. Vertailussa tarkasteltiin optimaalisten kyselyjen tehokkuutta ja raken-

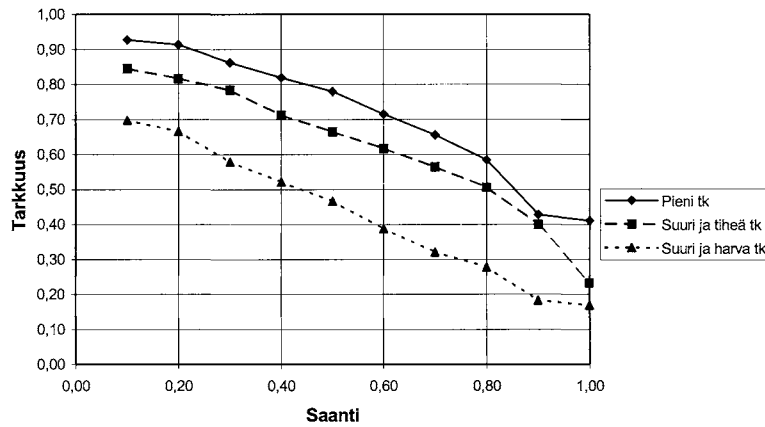
teita (tyhjentävyys ja kattavuus). Tutkimushypoteeseja muotoiltiin kaikkiaan 12, joista 6 liittyi suureen tarkkuuteen tähtääviin kyselyihin ja 6 suureen saantiin tähtääviin kyselyihin. (Sormunen 2000, 92-98).

Tutkimuksessa käytettiin Tampereen yliopiston informaatiotutkimuksen laitoksen TUTK-testikokoelmaa, joka koostuu 54.000 lehtiartikkelin tietokannasta, 35 testikysymyksen patterista ja näihin liittyvistä relevanssiarvioista. Olennainen osa testikokoelmaa ovat myös kattavat hakusuunnitelmat, jotka sisältävät 134 tunnistettua hakufasettia (2-5 fasettia per hakutehtävä) ja 2330 hakusanaa (1-74 sanaa per fasetti). (Sormunen 2000, 59-72).

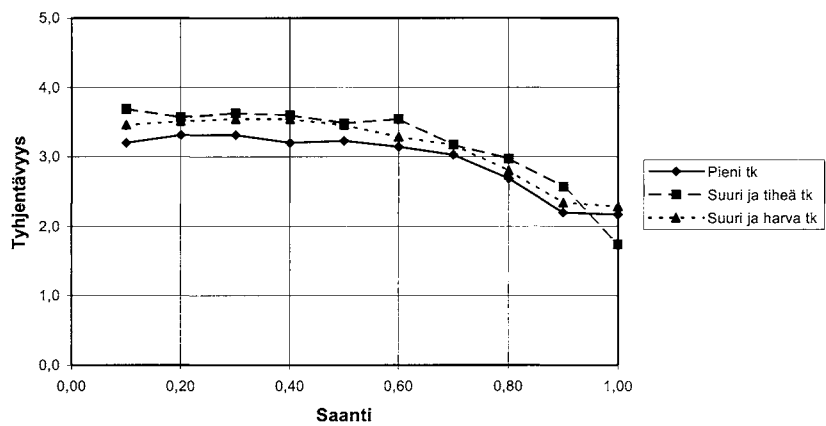
Tutkimuksessa pystyttiin löytämään uutta tietoa mm. parhaiten eri hakutilanteissa toimivista kyselyrakenteista sekä näyttämään missä tilanteissa perinteiset Boolean kyselyt toimivat hyvin, missä tilanteissa taas kohdataan ongelmia. Kuvissa 4-6 on esitetty esimerkkitulokset optimaalisten kyselyiden tarkkuuden, tyhjentävyyden ja suhteellisen kattavuuden vaihtelusta eri saantitasoilla kolmessa erityyppisessä tietokannassa. Tulokset näyttävät mm., että tarkkuus jäi molemmissa suurissa tietokannoissa pientä tietokantaa pienemmäksi. Suuri ja tiheä tietokanta, jossa relevanttien dokumenttien tiheys oli sama kuin pienessä tietokannassa, keskimääräinen ero oli vain noin 13%. Ns. suuressa ja harvassa tietokannassa (relevantteja sama määrä per kysymys kuin pienessä tietokannassa) tarkkuus jäi noin 40% pientä tietokantaa heikommaksi. Vaikka kyselyiden tyhjentävyys oli suurissa tietokannoissa korkeampi ei hakutulosten tarkkuutta pystytty ylläpitämään. Suuressa ja tiheässä käytettiin suhteellisesti enemmän hakusanoja per fasetti. Mitä useampia dokumentteja on löydettävä, sitä useampia rinnakkaisia ilmauksia on kyselyyn lisättävä.

Korkeimmalla saantitasolla on havaittavissa suuren ja tiheän tietokannan tarkkuus- ja tyhjentävyysarvoissa hyvin kiinnostava romahdus. Tässä tietokannassahan oli löydettävä noin viisinkertainen määrä relevantteja dokumentteja 100%:n saantiin yltämiseksi. Relevanttien dokumenttien tekstianalyysi paljasti, että syynä tyhjentävyyden ja samalla tarkkuuden romahtamiseen oli se, että joissakin relevanteissa dokumenteissa ei yksinkertaisesti esiintynyt hakukelpoisia ilmauksia läheskään kaikista hakusuunnitelman faseteista (käsitteistä). AND-operaattoria ei sen vuoksi voitu soveltaa ja tyhjentävyys putosi useissa hakutehtävissä yhteen (yhden fasetin kyselyt).

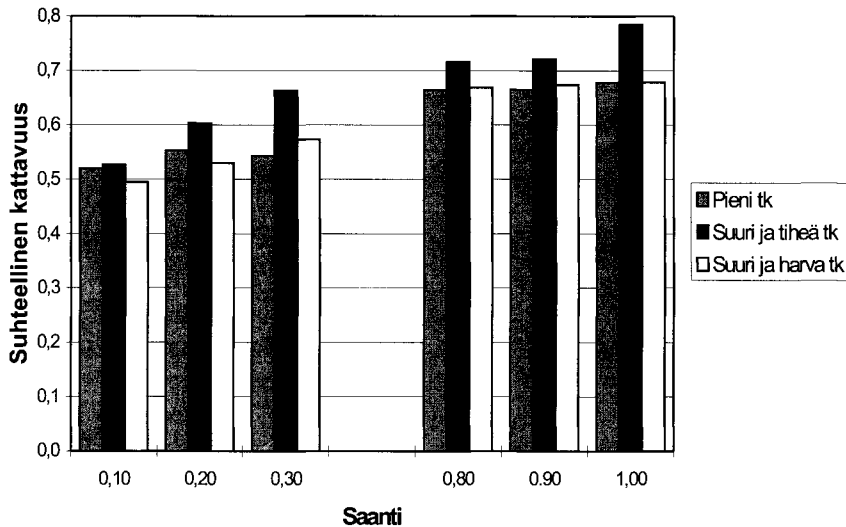
Edellä on esitetty vain näyte varsinkin laajan esimerkkievaluoinnin tuloksista. Työssä selvitettiin myös suureen tarkkuuteen tähtääviin kyselyihin liittyviä ilmiöitä. Osoitettiin mm. virheelliseksi tai ainakin puutteelliseksi aiemmat läheisyysoperaattoritutkimukset näyttämällä, että suureen tarkkuuteen tähtäävässä vapaatekstihaussa AND-



**Kuva 4.** Optimaalisten kyselyjen keskimääräinen tarkkuus eri saantitasoilla pienessä, suuressa ja tiheässä ja suuressa ja harvassa tietokannassa (35 hakutehtävää).



**Kuva 5.** Optimoitujen kyselyiden keskimääräinen tyhjentävyys eri saantitasoilla pienessä, suuressa ja tiheässä sekä suuressa ja harvassa tietokannassa (35 hakutehtävää).



*Kuva 6. Optimaalisten kyselyjen suhteellinen kattavuus pienessä, suuressa ja tiheässä sekä suuressa ja harvassa tietokannassa.*

operaattori toimii keskimäärin samalla tarkkuudella kuin läheisyysoperaattorit. Kyselyn tyhjentävyyttä on vain kasvatettava. Lisäksi havainnollistettiin, että lisäämällä Boolean kyselyjen tyhjentävyyttä voidaan parantaa kaikkein relevantimpien dokumenttien löytymistä. Tätäkään ilmiötä ei aiempi tutkimus ole havainnut. (ks. Sormunen 2000, 101-151).

## 5 Loppuarvio

Tutkimuksen keskeinen tieteellinen hyöty on siinä, että se kaataa raja-aitoja laboratorio- ja käyttäjäsuuntautuneiden tutkimuslinjojen väliltä. Tiedonhauksen ilmiöitä voidaan tutkia laboratorioympäristössä niin, että hakija kyselyjen muotoilun asiantuntijana otetaan mukaan tiedonhakuprosessiin eikä eristetä siitä. Toinen tärkeä metodologinen avaus on Harterin retrospektiivisen, eli relevanssietojen hyödyntävän menetelmäidean jalosta-

minen käytännössä sovellettavaksi evaluointimenetelmäksi.

Käytännön hakijan näkökulmasta tutkimuksessa pystyttiin hahmottelemaan perinteisten Boolean hakujärjestelmien maksimaalista toimintamekanismia suurissa tekstitietokannoissa. Tulokset auttavat tunnistamaan eri hakutilanteissa, mistä suunnasta parhaiten toimivaa Boolean kyselyn rakenne mahdollisesti löytyy.

Retrospektiivista lähestymistapaa on sovellettu varsin vähän tiedonhaku tutkimuksessa. Shaw (1995) sovelsi sitä probabilististen tiedonhaku järjestelmän optimaalisten parametrien määrittelyyn tavalla, joka antoi Robertsonille (1996) perustellut syyt saattaa tutkimuksen tulokset vähintäänkin kyseenalaisiksi ylisovitukseen (overfitting) perustuvan validisuuskritiikin perusteella. Kritiikin ydin oli siinä, että Shaw'n tapa soveltaa optimointia johti satunnaisten tilastollisten ilmiöiden (kuten kirjoitusvirheet) korostumiseen tuloksissa. Tulosten perustella ei voitu tehdä mitään ennusteita siitä, miten järjestelmän para-

metrit tulisi valita, että ne toimisivat optimaalisesti jossain toisessa kokoelmassa. Shaw'n epäonnisen tutkimuksen jälkeen retrospektiivistä menetelmää ei tiettävästi ole sovellettu ennen tätä tutkimusta. Retrospektiivinen lähestymistapa on kuitenkin kiinnostava ja perusteiltaan täysin pitävä kunhan sitä käytetään sille sopivien tutkimusongelmien ratkaisuun ja vältetään ylisovituksen ongelmat.

Väitöskirjatutkimuksessa kehitettiin optimointialgoritmi kyselyjen optimointiin. Kiinnostavalta jatkotutkimukselta vaikuttaa optimointialgoritmin korvaaminen tiedonhakupelillä ja asiantuntevilla hakijoilla. Tiedonhakupelin käyttö tarjoaa yksinkertaisen mahdollisuuden myös eri tiedonhakumallien väliselle vertailulle.

Hyväksytty julkaistavaksi 1.9.2000.

## Lähdeluettelo

- Frants, V.I., Shapiro, J., et al. (1999). Boolean Search: Current State and Perspectives. *J. Am. Soc. Inf. Sci* 50(1), 86-95.
- Harter, S.P. (1990). Search Term Combinations and Retrieval Overlap: A Proposed Methodology and Case Study. *J. Am. Soc. Inf. Sci* 41(2), 132-146.
- Hersh, W.R. & Hickam, D.H. (1995). An Evaluation of Interactive Boolean and Natural Language Searching with Online Medical Textbook. *J. Am. Soc. Inf. Sci* 48(7), 478-489.
- Ingwersen, P. & Willett, P. (1995). An Introduction to Algorithmic and Cognitive Approaches for Information Retrieval. *Libri* 45(), 160-177.
- Newell, A. (1968). Heuristic programming: Ill-structured problems. In: Arofonsky, J. (Ed.). *Progress in Operations Research*, Vol III, 360-414. New York.
- Paris, L.A.H. & Tibbo, H.R. (1998). Freestyle vs. Boolean: A comparison of partial and exact match retrieval systems. *Inf. Proc. Man.* 34(2/3), 175-190.
- Robertson, S.E. (1996). Letter to the Editor. *Inf. Proc. Man.* 32(5), 635-636.
- Saracevic, T. (1995). Evaluation of evaluation in information retrieval. In: Fox, E.A. et al. (Eds.), *SIGIR '95 - Proc. of the 18th Annual International ACM SIGIR Conference*. Washington July 9 - 13, 1995, p. 138-146.
- Shaw, W.M. (1995). Term-relevance computations and perfect retrieval performance. *Inf. Proc. Man.* 31(4), 491-498.
- Sormunen, E. (2000). A Method for measuring Wide Range Performance of Boolean Queries in Full-Text Databases. Doctoral Thesis. Tampere: University of Tampere. *Acta Electronica Universitatis Tamperensis*, ISBN: 951-44-4732-8, 231 p. URL: <http://acta.uta.fi/pdf/951-44-4732-8.pdf>.
- Tague-Sutcliffe, J. (1992). The pragmatics of information retrieval experimentation, revisited. *Inf. Proc. Man.* 28(4), 467-490.