

Informaatiotutkimuksen päivät 2010
21. - 22. lokakuuta, Tampere

ABSTRAKTI

Heikki Keskustalo

Vuorovaikutuksen simuloinnista tiedonhaussa

Heikki Keskustalo, Informaatiotutkimuksen ja interaktiivisen median laitos, Tampereen yliopisto, ccheke@uta.fi

Tiedonhakumenetelmien tehokkuutta voidaan tutkia yhdessä ääritapauksessa tilanteessa, jossa todelliset hakijat suorittavat omiin työtehtäviinsä liittyviä hakuja. Kuitenkin perinteisesti tehokkuutta on usein vertailtu kokonaan ilman käyttäjien läsnäoloa, testikokoelmien avulla. Testikokoelmat koostuvat kolmesta osasta – joukosta ennalta laadittuja kysymyksiä eli hakuaiheita, joukosta haettavia dokumentteja (tietokanta) ja korrektien vastausten joukosta (relevanssiarviot). Tiedetään siis ennalta, mitkä dokumentit hakumenetelmien tai järjestelmien tulisi löytää. Viime aikoina on toistuvasti raportoitu, että testikokoelmien avulla havaittu hakujärjestelmien keskinäinen paremmuus ei ole välttämättä merkityksellistä todellisessa hakutilanteessa. Todellisuudessa hakija voi siis menestyä yhtä hyvin kahdella eri järjestelmällä, vaikka toinen järjestelmä on testikokoelmapohjaisessa testauksessa osoittautunut toista paremmaksi (Smith & Kantor, 2008; Turpin & Hersh, 2001; Turpin & Scholer, 2006).

Perinteisessä testikokoelmapohjaisessa hakumenetelmien vertailussa hakutehokkuutta tarkastellaan suorittamalla kullakin hakumenetelmällä yksittäinen kysely useista eri hakuaiheista. Lopuksi eri menetelmien antamien tulosten laatua vertaillaan. Olennaista tässä perinteisessä mallissa on se, että hakijan piirteitä ja vuorovaikutusta ei mallineta. Hakumenetelmien tehokkuuden vertailu perustuu siis yhden ainoan kyselyn käyttöön kullekin hakuaiheelle (Järvelin & al., 2008). Tässä tilanteessa hyvin lyhyiden - esimerkiksi yksisanaisten - kyselyiden käyttö ei välttämättä ole kovinkaan mielekäästä. Tuloksen laadun arviointi perustuu puolestaan haettujen dokumenttien aiheenmukaiseen relevanssiin – eli siihen, kuinka paljon dokumentit sisältävät hakuaihetta koskevaa informaatiota (Sormunen, 2002). Lisäksi käytetyt tuloksellisuusmittarit perustuvat yksinomaan siihen, että tarkastellaan haettujen dokumenttien listoja (Azzopardi, 2007). Hakua ei siis nähdä useista yrityksistä koostuvana prosessina eikä prosessia oteta evaluoinnissa huomioon.

Todellisten hakijoiden toiminta eroaa merkittävästi edellä kuvatusta perinteisestä mallista (Kekäläinen & Järvelin, 2002). Hakijat suosivat tyypillisesti vuorovaikutusta hakujärjestelmän kanssa (Järvelin & al., 2008). He hakevat dokumentteja perustaen toimintansa vuorovaikutukselle - yrityksen ja erehdyksen periaatteelle. Todellisten

hakijoiden kyselyt ovat tyypillisesti hyvin lyhyitä (Jansen & al., 2000; Stenmark, 2008; Vakkari, 2000) – he eivät tahdo ilmaista kaikkea tietämäänsä, jos sen voi välttää. Lisäksi hakijat selailevat mielellään vain lyhyitä dokumenttistöjä (ehkä kymmenkunta dokumenttia) jos kysely epäonnistui. Hakijat ovat siis taipuvaisia yrittämään uudelleen sen sijaan että he jatkaisivat tulosjoukon selailua yhä pidemmälle. Hakijat myös usein tyytyvät muutamaan hyvään dokumenttiin. Lisäksi hakijat kiinnittävät huomiota hakuprosessiin kokonaisuudessaan – esimerkiksi pyrkien minimoimaan hakulausekkeiden kirjoittamisen vaivaa.

Voidaankin todeta, että edellä kuvatut perinteisen testikokoelmapohjaisen tiedonhaun oletukset ja todellisten hakijoiden käyttäytymispiirteet ovat monessa mielessä täysin vastakkaisia. Esimerkiksi hakija, joka kirjoittaa pitkän hakulausekkeen ja selailee sen jälkeen ensimmäiset 1000 haettua dokumenttia – tilanne, jota perinteinen laboratoriomalli saattaa mallintaa – on epäilemättä harvinaisuus.

Simulaatiot tarjoavat erään keinon vuorovaikutusprosessien piirteiden huomioimiseen testikokoelmapohjaisessa tiedonhaun tutkimuksessa. Simulaatiolla tarkoitetaan kokeiden tekoa perustuen malliin, joka on todellisuuden yksinkertaistettu kuvaus. Laajasti tulkittuina myös perinteiset testikokoelmapohjaisen tiedonhaun tutkimukset voidaan mieltää yksinkertaisiksi simulaatioiksi. Huomiota tulisikin kiinnittää siihen, millaiseen tiedonhaun malliin simulaatio perustuu ja mitä piirteitä mallissa tulisi olla. Esimerkkeinä aiemmista simulaatiotutkimuksista voidaan mainita Ahlgrenin (2004) tutkimus, jossa simuloidaan käyttäjäloukuttavuutta (mutta ei vuorovaikutusta) mallintamalla käyttäjiä, joiden kärsivällisyys vaihtelee, ja jotka antavat erilaisia arvoja eri relevanssitasoille kuuluville dokumenteille. White et al. (2005) on esimerkki simulaatiosta, jossa vuorovaikutusta mallinnetaan simuloimalla käyttäjää, joka tarkastelee dokumenttien rakennetta (kuten otsikoita tai parhaita lauseita) erilaisissa järjestyksissä samalla kun tutkitaan automaattisten relevanssipalautemekanismin kykyä parantaa tiedonhaun tuloksellisuutta näissä erilaisissa tilanteissa.

Informaatiotutkimuksen ja interaktiivisen median laitoksella on viime vuosina kehitetty tiedonhaun testikokoelmaperustaisen simulaation tutkimusta, jossa kiinnitetään huomiota käyttäjävuorovaikutuksen mallintamiseen. Toistaiseksi tutkimuksissa on keskitytty pintatason vuorovaikutuksen mallintamiseen. Tutkimuksissa on simuloitu kahta eri vuorovaikutustyyppiä, relevanssipalautetta ja hakulausekkeiden suoraa muokkausta. Relevanssipalautesimulaatioissa hakulausekettä muokataan automaattisesti, kun oletetaan että hakija tunnistaa relevantteja dokumentteja, joista poimitaan automaattisesti uusia hakusanoja (Keskustalo & al., 2008). Hakulausekkeiden suoran muokkauksen simulaatiossa mallinnetaan tilannetta, jossa hakija voi yrittää useita systemaattisesti uudelleenmuotoiltuja kyselyitä samasta hakuaiheesta (Keskustalo & al., 2009). Jälkimmäisissä simulaatioissa havaittiin, että erittäin lyhyiden kyselyiden käyttö kyselysekvensseinä voi olla rationaalista toimintaa vaikka yksittäin tarkasteltuina ne vaikuttavat erittäin huonolta vaihtoehdolta. Tämäntyyppisten simulaatioiden evaluointivaiheessa käyttäjänäkökulma korostuu. On kiinnitettävä huomiota siihen, mitä dokumentteja simuloitu hakija havaitsee, millaisia arvoja hän näille antaa ja millaista lopputulosta hän pitää menestyksellisenä.

Käyttäjävoroaikutussimulaatiot ovat lupaava menetelmä laajentaa perinteisiä hakukokoelmatestejä. Ne voivat auttaa tekemään ymmärrettäväksi todellisuudessa havaittua hakukäyttäytymistä. Ne myös tarjoavat mahdollisuuksia löytää vuoroaikutustapoja, jotka palvelevat hakijaa määriteltyjen rajoitteiden ja vaatimusten vallitessa. Vuoroaikutussimulaatioiden yhtenä tulevaisuuden haasteena on tarkastella hakujärjestelmien tehokkuutta ottaen huomioon sekä hakuprosessin olennaiset piirteet että tietynlaiselle hakijalle keskeiset menestyskriteerit.

Lähteet

- Ahlgren, P. (2004) The effect of indexing strategy-query term combination on retrieval effectiveness in a Swedish full text database. Dissertation. Valfrid, Sweden, 2004.
- Azzopardi, L. (2007) Position Paper: Towards Evaluating the User Experience of Interactive Information Access Systems. In: SIGIR'07 Web Information-Seeking and Interaction Workshop, 5 p.
- Jansen, M. B. J., Spink, A., Saracevic, T. (2000) Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web, *Information Processing and Management* 36(2), pp. 207-227.
- Järvelin, K., Price, S.L., Delcambre, L. M. L., and Nielsen, M. L. (2008) Discounted Cumulated Gain Based Evaluation of Multiple-Query IR Sessions. In: ECIR'08, pp. 4-15.
- Kekäläinen, J., Järvelin, K. (2002) Evaluating Information Retrieval Systems under the Challenges of Interaction and Multidimensional Dynamic Relevance. In: Proceedings of the 4th CoLIS Conference, pp. 253-270.
- Keskustalo, H., Järvelin, K., Pirkola, A. (2008) Evaluating the effectiveness of relevance feedback based on a user simulation model: effects of a user scenario on cumulated gain value. *Information Retrieval*, 11, pp. 209-228.
- Keskustalo, H., Järvelin, K., Pirkola, A., Sharma, T., Nielsen, M. L. (2009) Test Collection-Based IR Evaluation Needs Extension Toward Sessions – A Case of Extremely Short Queries. In: AIRS'09, pp. 63-74.
- Smith, C. L. and Kantor, P. B. (2008) User Adaptation: Good Results from Poor Systems. In: SIGIR'08, pp. 147-154.
- Stenmark, D. (2008) Identifying Clusters of User Behavior in Intranet Search Engine Log Files. *Journal of the American Society for Information Science and Technology*, 59(14), pp. 2232-2243.
- Sormunen, E. (2002) Liberal relevance criteria of TREC – Counting on negligible documents? In: SIGIR'02, pp. 320 – 330.
- Turpin, A. and Hersh, W. (2001) Why Batch and User Evaluations Do Not Give the Same Results. In: SIGIR'01, pp. 225-231.
- Turpin, A. and Scholer, F. (2006) User Performance versus Precision Measures for Simple Search Tasks. In: SIGIR'06, pp. 11-18.
- Vakkari, P. (2000) Cognition and changes of search terms and tactics during task performance: a longitudinal study. In: Proceedings of the RIAO 2000 Conference, Paris: C.I.D., pp. 894-907.

White, R. W., Ruthven, I., Jose, J. M., van Rijsbergen, C. J. (2005) Evaluating Implicit Feedback Models Using Searcher Simulations. *ACM Transactions on Information Systems*, Vol. 23, No. 3, 2005, pp. 325-361.