

Informaatiotutkimuksen päivät 2010
21. - 22. lokakuuta, Tampere

ABSTRAKTI

Sanna Kumpulainen

Bioteknologien tehtävälähtöinen tiedonhankinta : järjestelmien integroitu käyttö työtehtävissä

Sanna Kumpulainen, Tampereen yliopisto, INFIM, sanna.kumpulainen@uta.fi.

Johdanto

Molekyyli lääketiede tarjoaa hyvän tutkimuskohteen tehtävälähtöiselle tiedonhankinnalle monimutkaisten työtehtävien ja hajautetun informaatioympäristön vuoksi. Tämä tutkimus keskittyy bioteknologien todellisiin tehtäväprosesseihin ja eri tietokanavien integroituun käyttöön työtehtäviä suoritettaessa. Tutkimus perustuu monipuoliseen havainnointiaineistoon. Tutkimuksessa tarkastellaan työtehtäviin liittyvää tiedonhankintaa kolmella eri tehtäväkompleksisuustasolla, eri järjestelmien integroitua käyttöä sekä kyselytyyppejä näillä tasoilla. Tämä tiivistelmä perustuu aiemmin julkaistuun laajempaan tutkimukseen. (Kumpulainen & Järvelin 2010).

Tiedonhaun web-käyttäjätutkimuksissa käytetään usein yhden hakujärjestelmän käytöstä kertovia lokeja (He & Göker 2000). Tutkimusten tulokset saattavat olla siirrettävissä koskemaan muita hakukoneita, mutta erilaisten järjestelmien yhteiskäyttöä ei voida näin saada selville. Pelkkiin palvelinlokiaineistoihin on vaikea liittää kontekstitytöä, jolloin päätelmien teko käyttäjien tarpeista ja päämääristä voi olla haastavaa. Epämääräiset ja monimutkaiset tiedontarpeet saattavat jäädä tunnistamatta.

Tehtävälähtöistä tiedonhankintaa on jonkin verran tutkittu, mutta tarkasteltavat tehtävät ovat olleet tiedonhakutehtäviä. (Kellar, Watters & Shepherd 2007; Lin & Wilbur 2009). Myös biolääketieteen työtehtäviä on tarkasteltu, mutta nämä tutkimukset keskittyvät joko yksittäiseen tehtävään tai teoreettisiin tehtäväluokkiin. (Bartlett & Toms 2005; Stevens 2001).

Tutkimusasetelma

Kuutta tutkijaa kahdesta eri tutkimusryhmästä havainnoitiin keskimäärin 24 tuntia kolmesta kahdeksaan viikkoa. Tutkijat tekivät päivittäistä työtään omassa työympäristössään. Aineisto koostuu hakulokeista, haastatteluista ja havainnointidatasta, jotka kerättiin puolen vuoden aikana vuosina 2007–2008. Haastattelut tehtiin ennen havainnointia, ja niiden tarkoitus oli lähinnä taustoittava.

Tehtäväsessiot (24 sessiota) tunnistettiin ja havainnointimuistiinpanot ja lokit yhdistettiin sessioittain. Sessioiden pituus vaihteli 30:stä 170 minuuttiin. Lokista saatiin selainliikenne, kun taas PC:n käyttö, painetut ja henkilölähteet selvitettiin havainnointidatasta. Käytetyt

tietoresurssit luokiteltiin kanaviksi. Kaikki siirtymät eri kanavien välillä laskettiin. Sessiot jaettiin kolmeen eri kompleksisuusluokkaan: vaikeisiin, keskivaikeisiin ja rutiinisessioihin. Kompleksisuus määrittyi sen mukaan, oliko bioteknologueilla käsitystä tehtävän suorittamiseen tarvittavista resursseista, tehtäväprosessista tai lopputuloksesta. Kyselyt luokiteltiin käytetyn tietokannan sekä tiedontarpeen mukaan.

Tulokset

Sessioissa käytettiin useita erityyppisiä resursseja, kanavia, yhdessä. Tyypillisintä oli siirtää tietoa PC:n ja biotietokantojen välillä. Esim. laboratoriodataa käytettiin syötteenä haettaessa biotietokannoista, tai biotietokannoista haettiin dataa PC:lle muokattavaksi jatkohakuja varten. Kirjallisuutta haettiin kirjoitustehtävissä lähdeviitteitä varten. Suosittua oli kopioida lauseita käsikirjoituksesta ja kopioida näitä joko hakukoneeseen tai kirjallisuusviitetietokantaan (PubMed).

Tarkasteltaessa siirtymiä sessioiden kompleksisuuden mukaan, vaikeissa sessioissa siirtymät hajaantuvat eri kanavien välille, kun taas keskivaikeissa ja rutiinisessioissa siirtymät keskittyvät PC:n ja biotietokantojen välille (PC:ltä biotietokantoihin 32,6 % ja toisinpäin 32,0 %).

Sessioiden kyselyistä (n=289) suurin osa tehtiin biotietokantoihin (61 %). Kyselyt jakaantuvat eri tavoin eri kompleksisuustasoilla. Suurin osa vaikeiden sessioiden kyselyistä oli faktahakuja biotietokantaan. Keskivaikeissa yleisin kyselytyyppi oli aihehaku biotietokantaan ja rutiinitasolla on datan keräämistä jatkoanalyysijä varten sekä kirjallisuuden keräämistä tietokantapäivityksiä varten. Tämä osoittaa, että vaikeissa tehtävissä tiedontarpeet ovat niin epämääräisiä, että aihehakuja ei voida tehdä, sillä aihetta ei tunneta. Web-kyselyistä 53,4 % oli navigointi-, neljännes oli fakta- ja viidennes aihehakuja. Kaiken kaikkiaan webhakuja oli 14,2 %, joten nämä muodostivat tiedonhausta varsin pienen osan.

Johtopäätökset

Mitä vaikeampi sessio on, sitä integroidumpaa ja vaihtelevampaa on eri kanavien käyttö eli rutiinitehtävät keskittyvät kahden kanavan välille kun taas vaikeissa sessioissa käytetään useita eri kanavia vaihdellen.

Useita eri tietoresursseja integroidaan aktiivisesti työtehtäviä suoritettaessa. Integrointi voi olla joko automaattista, puoliautomaattista tai sitten se tehdään käsin. Automaatio tarkoittaa implisiittistä informaation keräämistä ja koostamista eri lähteistä. Puoliautomaattisessa integroinnissa eri tietokantojen välillä on linkkejä, mutta käyttäjän täytyy kuitenkin painaa linkkiä, kun taas käsin tehtävä integrointi tarkoittaa sitä, että käyttäjän täytyy tehdä ensin haku tietokantaan saadakseen syötettä toiseen järjestelmään, ja mahdollisesti sitä on vielä muokattava PC:llä siinä välissä. Työelämän tehtävänsuoritusseesiot ovat täynnä näitä integroinnin eri asteita täysin automaattisesta manuaaliin, käsin tehtävään integrointiin. Työtehtävien tekemistä voitaisiin helpottaa nostamalla automaation astetta (a) lisäämällä automaattista, koostavaa tiedonhakuja eri järjestelmistä, (b) resursseja linkittämällä, (c) tarjoamalla tehtävän kannalta relevantteja hakuavaimia sekä (d) harmonisoimalla datan semanttista epäyhätenäisyyttä eri tietokannoissa.

Lähteet

Bartlett, J.C. & Toms, E.G. (2005). Developing a protocol for bioinformatics analysis: an integrated information behavior and task analysis approach. *Journal of the American Society for Information Science and Technology*, 56(5):469–482.

- He, D., & Göker, A. (2000). Detecting session boundaries from Web user logs. Teoksessa: Proceedings of the BCS/IRSG 22nd Annual Colloquium on Information Retrieval Research, s. 57–66. Cambridge, UK.
- Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology*, 58(7):999–1018.
- Kumpulainen, S. and Järvelin, K. (2010). Information interaction in molecular medicine: integrated use of multiple channels. Teoksessa: IiiX '10: Proceeding of the third symposium on Information interaction in context, s. 95-104. New York, NY, USA. ACM.
- Lin, J. and Wilbur, W. J. (2009). Modeling actions of PubMed users with n-gram language models. *Information Retrieval*, 12(4):487–503.
- Stevens, R., Goble, C., Baker, P. & Brass, A. (2001). A classification of tasks in bioinformatics. *Bioinformatics*, 17(2):180–188.