

*Sanna Kumpulainen**

Tehtäväperusteinen tiedonhaku molekyylilääketieteessä

Johdanto

Ihmisen olemassaoloa määrittää osittain käsitys hakemisesta inhimillisenä toimintana. Etsimme ja jahtaamme ruokaa ja suojaa, sekä aineettomia asioita kuten kokemuksia ja oikeutta. Eräs tällainen aineeton hakemisen kohde on tieto. Tiedonhaku voidaan nähdä ihmiselle tyypillisenä toimintana, joka rinnastuu oppimiseen ja ongelmanratkaisuun. Nämä kaksi ominaisuutta, oppimis- ja ongelmanratkaisutaidot, ovat olleet keskeisiä ihmisen selviytymiselle kautta aikojen. Näitä ominaisuuksia käytetään nykyisessä tietoyhteiskunnassa päivittäin mm. erilaisissa työtehtävissä. Tehtävien tavoitteiden saavuttamiseksi tarvitaan tehtäväspesifiä tietoa. Tietoa voidaan joko tuottaa itse, tieto voidaan muistaa tai hakea. Tietoa voidaan hakea kysymällä lähimmältä henkilöltä, lukea jostain saatavilla olevasta muistiosta tai muusta paperista, tai sitä voidaan hakea erilaisista sähköisistä tiedonhaku-järjestelmistä.

Tässä tutkimuksessa ollaan kiinnostuneita lähinnä jälkimmäisestä, eli tiedonhausta tiedonhaku-järjestelmistä. Väitöskirjassani käsitellään tiedonhakua molekyylilääketieteen työtehtävien kontekstissa. Molekyylilääketiede on erittäin tietointensiivinen ala, ja täten se tarjoaa hyvän tutkimuskohteen tehtäväperusteisen tiedonhaun tutkimukselle. Molekyylilääketieteessä perehdytään sairauksien syntymekanismiin, diagnostiikkaan sekä hoitamiseen molekyylili- ja solutasolla.

*Sanna Kumpulaisen väitöskirja *Task-based information access in molecular medicine: task performance, barriers and searching within a heterogeneous information environment* (Tehtäväperusteinen tiedonhaku molekyylilääketieteessä: tehtävien suorittaminen, esteet ja hakeminen epäyhtenäisessä informaatioympäristössä) tarkastettiin Tampereen yliopistossa 5.12.2013.

Väitöskirja kuuluu informaatiotutkimuksen ja interaktiivisen median oppialaan, joka on luonteeltaan monitieteinen. Siinä tutkitaan monimuotoisia informaatioon ja sen käyttöön liittyviä ilmiöitä. Tieto- ja informaatiokäyt-
tämisen tutkimuksen osa-alueella on ollut kaksi vahvaa perinnettä: tiedonhankinnan tutkimus, joka ammentaa osin sosiologisesta tutkimusperinteestä, sekä tiedonhaun tutkimus, joka on enemmän järjestelmiin keskittyntynyt ja vahvasti tietojenkäsittelyyn päin kallellaan. Tiedonhankinnan tutkimuksessa on tutkittu esimerkiksi erilaisten ryhmien tiedontarpeita ja tiedonhaun kanavia. Tiedonhaku-tutkimuksessa on perinteisesti ollut vahvoilla erilaisten hakujärjestelmien kehittäminen ja niiden evaluointi. Kummatkin tutkimusperinteet kuitenkin tarkastelevat perimmiltään samaa ilmiötä, tiedon äärelle hakeutumista. Tässä väitöstutkimuksessa on pyritty ottamaan molempia näkökulmia huomioon: tarkastelun keskiössä on tiedonhaku ja tietojärjestelmien käyttö, mutta käyttöön on pyritty myös tuomaan toiminnan taustoja tarkastelemalla tätä tietojärjestelmien kautta tapahtuvaa tiedonhakua työtehtävän suorituksen kautta. Informaatiovuorovaikutusta tarkastellaan siis tehtäväperusteisesti.

Informaatiovuorovaikutus on yksi kuumista sanoista tämän hetken tiedonhaku-tutkimuksessa ja kuvaa hyvin tämän tutkimuksen lähestymistapaa. Sillä tarkoitetaan karkeasti sanottuna sitä, miten tiedonhakijat ovat vuorovaikutuksessa informaation kanssa, järjestelmistä huolimatta. Järjestelmät nähdään ikään kuin välttämättömänä pahana, vain välineinä informaatioon kiinni pääsemiseksi. Järjestelmät ovat siis tarpeellisia, mutta käyttäjää ei välttämättä kiinnosta järjestelmän käyttö, vaan sen kantama tieto. Itse järjestelmän kanssa tapahtuvaa vuorovaikutusta tarkastellaan vain välillisesti, eikä järjestelmien piirteitä ole tässä varsinaisesti analysoitu.

Tässä tutkimuksessa on oikeastaan kolme keskeistä osaa, joita tarkastellaan: informaatioympäristö, vuorovaikutus sen kanssa eli se varsinainen tiedonhaku, sekä menetelmät, eli miten voisi parhaiten tutkia tehtäväperusteista tiedonhakua. Menetelmien osuus kietoutuu luontevasti sekä informaatioympäristön että vuorovaikutuksen tutkimukseen.

Aloitan informaatioympäristöstä. Informaatioympäristö tarkoittaa kaikkia mahdollisia tiedonlähteitä, joita käytetään työtehtävien suorituksessa. Tutkimusprosessi aloitettiin kartoittamalla millaisista tiedonlähteistä molekyyliiläketieteen informaatioympäristö koostuu. Jo tässä vaiheessa syntyi vahva ajatus siitä, että vuorovaikutusta informaatioympäristön kanssa olisi mielekästä tutkia prosessina. Aika pian kävi selväksi, että kyselylomake sellaisenaan ei ole riittävä väline keräämään sopivaa tietoa. Päädyimme täydentämään kyselytutkimusta tekemällä haastatteluita, joiden aikana tutkijat saivat demonstroida tiedonhakutehtäviään ja järjestelmien käyttöä.

Tämä niin sanottu merkityksellisten kokemusten tekniikka palveli hyvin, kun haluttiin vain täydentää kyselyllä saatuja tietoja. Kuitenkin se jätti paljon epäselvyyksiä. Informantit valitsivat tyypillisiä tehtäviä demonstroitavaksi, mutta nämä olivat usein erittäin helppoja ja yksinkertaisia. Tässä on ristiriita sen kanssa, että informaatiotutkimuksessa ollaan ensisijaisesti kiinnostuneita palvelemaan nimenomaan haastavia, paljon informaatiota vaativia tehtäviä. He demonstroivat tehtäviä, jotka sisälsivät tyypillisiä hakuja ja hakustrategioita. Tällä tavoin ei esimerkiksi pystytä tutkimaan tiedonhaun esteitä koko kirjossaan, vaan ainoastaan tietyn tason ongelmat ovat havaittavissa. Tässä syntyi aito ongelmatilanne, eli miten sitten saisi selville ne oikeasti haasteelliset, kompleksit tehtävät näiden tyypillisten lisäksi, jotta voitaisiin tarkastella niiden tiedonhaun ilmiöitä.

Aineiston kerääminen varjostamalla

No mikä sitten olisi järkevä tapa tutkia todellisen maailman tilanteita? Vastaukseksi, ei mitenkään yllättävästi, muotoutui havainnointi: mennään paikalle katsomaan, kun työtä tehdään. Havainnointimenetelmäksi valikoitui varjostusmenetelmä, jossa seurataan varjostettavaa kohdetta pitkin päivää erilaisissa tilanteissa. Tässä

tutkimuksessa siis kuitenkin keskityttiin vain tiedonhakuintensiivisiin tilanteisiin, ja jätettiin palaverit ja laboratoriotyöskentely suurimmaksi osaksi väliin, vaikka kyllä varjostaja osaan näistäkin tunki mukaan. Varjostus on erittäin hyvä keino kerätä todellisen maailman tilanteisiin liittyviä aineistoja, mutta se on myös monin tavoin haastava: on vaikeaa päästä sisään organisaatioihin ja saada varjostettavat mukaan tutkimukseen. On aika suuri kynnyks päästää varjostaja hengittämään omaan niskaansa jopa kuukausiksi. Lisäksi varjostus on kognitiivisesti raskasta kummallekin osapuolelle. Varjostamalla kerätty aineisto on jäsentymätöntä ja se täytyy jälkikäteen järjestää. Tämä on työlästä, eikä välttämättä lainkaan miellyttävää. Tällä tavoin kuitenkin saatiin erittäin rikasta aineistoa, joka kattoi kaikenlaisia työtehtäviä joissa haettiin tietoja.

Varjostuksen aikana kerättiin myös lokitietoja internet-selaimen käytöstä, jolla saatiin tarkempaa kvantitatiivista aineistoa laadullisen varjostusaineiston tueksi. Lisäksi kerättiin lukuisia määrää valokuvia kaulassa kannettavalla kameralla. Nämä kuvat osoittautuvat hyväksi välineeksi tapahtumien muistiinpalauttamiseksi, vaikkei niitä varsinaisesti sitten käytettykään systemaattisessa analyysissä laatuselkkojen vuoksi.

Tällainen aineistotriangulaatio, eli erilaisten tutkimusaineistojen yhdistäminen, palvelee hyvin sekä tiedonhaun että sen kontekstin tutkimista. Saimme kerättyä monipuolisen aineiston kuuden tutkijan todellisista työtehtävistä, joten tiedonhakua pystyttiin analysoimaan tehtäväperusteisesti.

Tehtäväperusteisuus

Tehtäväperusteisuudessa lähtökohtana on työtehtävä, tässä tutkimuksessa molekyyliiläketieteen työtehtävä. Tehtäväperusteisessa tiedonhaussa on tarkoituksena tutkia tehtävän ja tiedonhaun prosessin yhteyksiä. Tehtävä on määrätietoista toimintaa, jolla on tavoite. Tavoitteen saavuttamiseksi saatetaan hakea tietoa.

Tässä tutkimuksessa käytettiin havainnointivien omia käsityksiä työtehtävistä, ja työtehtävät analysoitiin työntekoprosesseina. Näitä prosesseja kutsutaan sessioiksi/istunnoiksi. Nämä määritettiin siten, että jos tehtävän teko lopetettiin, loppui myös sessio. Samoin, jos työtehtävä vaihtui, päättyi yksi sessio ja toinen alkoi. Nämä sessiot analysoitiin niiden

kompleksisuuden mukaan. Kompleksisuuteen vaikutti havainnoitavan käsitykset työtehtävän tavoitteista, tarvittavista tietojärjestelmistä, sekä tehtävän prosessin tuntemus. Jos nämä pystyttiin selkeästi määrittelemään, katsottiin tehtävä yksinkertaiseksi, rutiininomaiseksi, jos taas näistä yksi tai kaikki olivat epäselviä, tehtävä määritettiin kompleksiseksi. Lisäksi määritettiin välimuoto, semi-kompleksinen tehtävä. Tiedonhakuja tarkasteltiin näissä eri kompleksisuusasteisissa tehtävissä.

Tehtävä – tiedontarve – tiedon hakeminen

Tehtävä nähdään toimintana, joka synnyttää tiedontarpeen. Tiedontarvetta on kuvattu epämääräisinä tuntemuksina sekä aukkona tietämyksessä. Tiedontarve voi olla hyvin jäsentynyt, tai sitten erittäin epämääräinen. Nämä kaksi ääripäätä johtavat erilaisiin tiedonhakuihin.

Hyvin jäsentyneessä tiedontarpeessa siihen vastaamiseen voi riittää yksinkertainen faktatieto, esimerkiksi puhelinnumero tai henkilön nimi. Epämääräinen tiedontarve taas on vaikeampaa muotoilla kysymysmuotoon. Monet hakujärjestelmät perustuvat ajatukseen, että käyttäjä kuvailee haluamansa sisällön mahdollisimman tarkasti hakusanoilla. Todellisuudessa on erittäin hankalaa kuvailla haluttuja tietoja, jos ei tiedetä mitä se voisi olla, minkälainen tieto vastaisi tiedontarpeeseen tai että onko sellaista ylipäänsä olemassa. Tämä usein johtaa erilaisiin selailu- ja hakupolkuihin, joiden avulla yritetään oppia jotain myös käsillä olevasta tehtävästä ja tiedontarpeesta. Tiedontarve saattaa jäsentyä, jolloin hakeminen tulee helpommaksi, voisi jopa sanoa mahdolliseksi.

Tiedonhaku työtehtävissä

Molekyylilääketieteen tutkijoiden tiedonhaku näyttäisi olevan varsin monimuotoista työtehtävien aikana. Yksi asia, mitä tämä tutkimus nostaa esille, on järjestelmien yhteiskäyttöisyys. Todellisissa tiedonhaun tilanteissa harvoin yksi tiedonlähde palvelee täysin tiedontarpeen täyttymistä. Jopa rutiininomaisissa työtehtävissä tiedonhauk näyttäytyivät monimutkaisina hakukuvioina. Tämä vaikuttaa johtuvan siitä, että erilaiset

tiedonhakujärjestelmät tarjoavat erilaisia lähestymistapoja ja näkökulmia kulloiseenkin tiedontarpeeseen, eikä yksittäinen järjestelmä pystynyt tarjoamaan koko tiedontarvetta kattavaa esitystä, vaan sitä haettiin palasina useista järjestelmistä. Joskus tilanne oli se, että sopiva hakuavain puuttui, ei ollut sellaista kahvaa jolla saada kiinni halutusta tiedosta. Tällöin jouduttiin ensin hakemaan sopivaa esitysmuotoa hakuavaimelle, jotta sitten voitaisiin hakea haluttu tieto toisesta järjestelmästä. Jos tehtävän monimutkaisuus lisääntyi, hakupolkujenkin muoto muuttui. Tällöin esimerkiksi haluttiin hakea samantyyppisiä kohteita ikään kuin listana, jolloin syötettiin hakukoneeseen yksittäin jokainen hakusana, ja kopioitiin tulokset yksitellen listaan.

Haastavimmissa komplekseissa tehtävissä saattoi esiintyä ns. marjanpoimintahakupolkuja: haettiin löyhiä, asiaan liittyviä näkökulmia erilaisista tietokannoista, ja yritettiin oppia tällä tavoin saaduista vihjeistä enemmän itse tehtävästä. Usein näissä on mukana epäselvä työprosessi: jos sitä tietoa ei ollut, miten tehtävä tulisi suorittaa, oli sellaisen tiedon suodattaminen saatavilla olevasta tiedosta ja datasta aika hankalaa. Proseduraalisen tiedon haku ei ole helppoa, pelkästään senkin vuoksi, että sitä ei ehkä ole olemassa yksiselitteisessä dokumentoidussa muodossa. Tämä saattoi aiheuttaa hämmennystä ja lisätä epävarmuutta tehtävän suoritukseen.

Yksi tässä tutkimuksessa tarkastelluista ilmiöistä olivat tiedonhaun esteet. Esteeksi katsottiin mikä tahansa tapahtuma, joka pysäytti tai merkittävästi hidasti työntekoa havainnoituissa sessioissa. Esimerkiksi edellä kuvattu listatyyppinen haku aiheutti loputonta kopiointia ja tiedon yhdistelyä. Usein tieto saattoi olla sellaisessa muodossa, että sitä ei saatu suoraan käyttöön, vaan se täytyi ensin muokata muilla välineillä sopivaksi. Näitä ylimääräisiä veivauksia työn kulussa kutsutaan kiertoteiksi, workarundeiksi. Jos järjestelmät eivät tukeneet työn logiikkaa, aiheutti se usein paljon ylimääräistä vaivaa. Tavallisin estetyyppi oli ihan vain teknologinen este tietojärjestelmän käytön kanssa: systeemit eivät toimineet siten kuin piti ja ne kaatuivat.

Implikaatiot

Tämä väitöstutkimus on deskriptiivinen, laadullinen tutkimus, jolla pyrittiin kuvaamaan

kuuden varjostetun tutkijan työssä tapahtuvaa tiedonhakua. Tässä tutkimuksessa ei alun perinkään pyritty suureen yleistettävyyteen, vaan tunnistamaan sellaisia tiedonhaun muotoja ja ilmiöitä, joilla on merkitystä tiedonhakijoille työtehtävien suorituksen näkökulmasta. Tällaisia kuvailevia tutkimuksia voidaan käyttää ponnistuslautoina selittäville tutkimuksille sekä informoimaan muita tutkimusasetelmia, kuten esimerkiksi laboratoriossa teetettäviä tiedonhaun käyttäjätutkimuksia. Näitä tutkimuksia voidaan informoida sellaisista tiedonhaun piirteistä, jotka olisi hyvä huomioida, mutta joita ei ole perinteisesti otettu mukaan tällaisiin tutkimusasetelmiin. Yleistettävyyttä saattaa ehkä olla heikko, sillä tutkimusaineisto oli suppea sekä sovellusala, molekyyli- ja lääketiede, erikoinen, mutta tutkimusmenetelmien osalta tämä tutkimus näyttäisi osoittavan, että tällainen todellisen maailman tutkiminen tuottaa erittäin rikkaa aineistoa, joka saattaa opettaa paljon siitä miten tiedonhakua tehdään oikeasti.

Lisäksi tällaisen tutkimuksen avulla voidaan selvittää, miten toimijat kokevat erilaiset tiedonhaun tilanteet. Se myös lisää tietämystä menetelmistä eli siitä, miten vuorovaikutteista tiedonhakua voisi tutkia. Se, mitä tästä voi oppia, saattaisi auttaa kokeellisen tutkimuksen tekemistä. Kaikki kokeelliset tutkimukset eivät ole välttämättä kytköksissä todellisen maailman ilmiöihin.

Lopuksi

Tiedon määrä lisääntyy jatkuvasti, eikä molekyyli- ja lääketieteen informaatioympäristö tee poikkeusta. Suurille tietoaineistoille on ominaista monimuotoisuus ja reaaliaikaisuus. Uusia biologisia aineistoja syntyy jatkuvasti, ja aineistomäärät ovat massiivisia. Isosta datamassasta voidaan löytää heikkoja, mutta olennaisia signaaleja. Samankaltaisuuksien ja säännönmukaisuuksien sijaan voi rikkaasta datasta etsiä myös poikkeamia, niitä mielenkiintoisia yksittäistapauksia, jotka tulee tavoittaa. Tällainen

tieto avaa mahdollisuuden oppia uutta ja nähdä suuntaa tapahtumassa oleville muutoksille. Sairauksien ennaltaehkäisy voi kehittyvän teknologian myötä nousta aivan uudelle tasolle.

Datamäärien kasvaessa kuitenkin myös niiden hallinnan vaatimukset kasvavat. Suuresta määrästä on vaikeampaa tavoittaa olennaisia asioita. Lisäksi tarvitaan parempia tapoja esittää ja yhdistellä sekä visualisoida monimuotoista ja epäyhtenäisiä datamassoja. Tulevaisuudessa ehkä tiedonhaussa on itsestään selvää, että tietoja tarjotaan automaattisesti eri järjestelmistä ilman, että tiedonhakijan tarvitsee sitä erikseen hakea tai pyytää. Jos tietojärjestelmien kehityspuolella olisi enemmän ymmärrystä tiedonhakijoiden käyttötilanteista ja siten tarpeista, ehkä järjestelmien kehitys voisi harpata ison askeleen eteenpäin. Päämääränä tulisi kuitenkin olla aina tiedonhakijan tavoitteiden täyttäminen, ja tämä on myös laadullinen ongelma.

Tiedonhakijoihin ja oletuksiin tiedonhaku-ikäytymisestä suhtaudutaan yhä samoin kuin kymmeniä vuosia sitten: tarkastellaan tehtyjä kyselyitä, ja pyritään vastaamaan kyselyihin sopivalla tulosjoukolla. Tämän tutkimuksen aineisto vihjailee, että erilaisia aineistoja yhdistellään tietoa hakiessa: esimerkiksi kuvia, biologista dataa ja tekstiaineistoja. Hakuteknologian tulisi kyetä hakemaan, suodattamaan, yhdistelemään, sulauttamaan ja prosessoimaan lukuisia ja hajautettuja tiedonlähteitä erikielisistä sisällöistä. Nykyinen vallitseva hakutapa on Googlehakutyypinen: pari sanaa kirjoitetaan pieneen laatikkoon. Eikö nykyinen teknologia mahdollista mitään muuta tapaa kuvata aukkoa tietämyksessä kuin tämä? Mitä jos tulevaisuudessa järjestelmien kanssa toimittaisiinkin jotenkin toisin? Sitä ei ehkä ratkaista pelkästään järjestämällä hakutulostoja ja lisäämällä tehoa, vaan siihen tarvitaan myös uusia lähestymistapoja. Siinä voisi osaltaan auttaa syvälinen ymmärrys hakijoiden tarpeista todellisissa tilanteissa.