

## 3.2 Tiedonhaun tutkimus

*Heikki Keskustalo, Paavo Arvola ja Sanna Kumpulainen*

- Tiedonhaun tutkimuksen keskeisenä tavoitteena on ymmärtää tiedonhakua ilmiönä.
- Tyypillisiä tutkimuskohteita ovat hakija, dokumentit, hakualgoritmit ja haettu tai muodostettu informaatio.
- Tiedonhaun tutkimustyyppit eroavat toisistaan keskeisesti siinä, millaisena *hakija* tutkimuksissa näyttäytyy. Tutkimuksessa läsnä oleva hakija voi olla autenttinen toimija tai koehenkilö suorittamassa annettua tehtävää – tai läsnä vain välillisesti.
- Hakijan läsnäolon tapa tutkimuksessa vaikuttaa siihen, millaisia kysymyksiä voidaan tutkia.
- Tiedon kenttä muuntuu alati tarjoten tutkimukselle jatkuvasti uusia kysymyksiä.

Tiedonhaun tutkimuksessa tarkastellaan, kuinka erilaiset tietojärjestelmät tuottavat relevanttia informaatiota hakijalle. Oppialana tiedonhaku voidaan sijoittaa informaatiotutkimuksen ja tietojenkäsittelytieteen rajamaastoon. Alan pioneerien Saltonin ja McGillin' mukaan tiedonhaussa ollaan kiinnostuneita *informaatioyksiköiden* (*information items*, ks. luku 2.3) esittämisestä, tallennuksesta, organisoinnista ja hakemisesta. Perinteisen dokumenttien haun rinnalle kasvavan mielenkiinnon kohteiksi ovat viime aikoina nousseet informaatiota *muodostavat* eli generoivat järjestelmät, kuten ChatGPT.<sup>2</sup> Haettaessa “valmiita” informaatioyksiköitä tai *muodostetta*ssa informaatiota dynaamisemmin voimme kummassakin tapauk-

<sup>1</sup> Salton & McGill 1987.

<sup>2</sup> Ks. esim. Wolfram 2023.

nessä puhua tiedonhausta. Tiedonhaun käyttöalat, tekniikat ja käytötavat muuntuvat alati ja esimerkiksi web-sovelluksia käytettäessä tiedonhaku sulautuu usein huomaamattomaksi osaksi erilaisia arjen palveluita.

Vaikka sana *tiedonhaku* viittaa vain tiedon hakemiseen, tutkimusalan nimityksenä termi kattaa laaja-alaisen joukon ilmiöitä. Tiedonhakua tutkittaessa hakijan tyypillisesti ajatellaan olevan kiinnostunut *relevantin* informaation löytämisestä tai muodostamisesta – ei dokumenttien hakemisesta sinällään. Tällöin dokumenttien ajatellaan *sisältävän* tulkittavissa olevaa informaatiota.

## Tiedonhaun tutkimusasetelmat

Tiedonhaun tutkimuksen yleisenä tavoitteena on ymmärtää tiedonhakua ilmiönä ja kehittää parempia hakujärjestelmiä ja -menetelmiä. Tämän kirjan luvussa 3.1 esitetään tiedonhaun yleinen malli, jonka pääelementit ovat hakija, dokumentit, hakualgoritmi sekä haun lopputulos eli haettu informaatio. Näiden elementtien avulla perinteinen tiedonhakuprosessi voidaan pelkistää seuraavasti: (i) hakija ilmaisee tiedontarpeensa hakujärjestelmälle käyttäen apuna kyselyä, (ii) järjestelmä antaa tarjolle dokumentteja haettaviksi sopivia ”tietorakenteita”, kuten hakemistoa, hyödyntäen, (iii) algoritmi vertailee tiedontarpeen ja dokumenttien ilmaisuja ja (iv) hakutulokset esitetään hakijalle, esimerkiksi linkkilistana.

Käytännössä tiedonhakijat ja hakutilanteet voivat olla hyvin monityyppisiä. Tiedonhaun *tutkimus* voidaan usein nähdä hakutilanteiden *simulaationa*, jossa tosiasiallista hakutilannetta jäljitellään ja tilanteen valikoituja piirteitä tarkastellaan lähemmin. Yksittäisessä tutkimuksessa käytetty hakijan ja hakuprosessin malli voi tällöin olla karkean yksinkertainen tai huomattavan hienojakoinen ja monimutkainen. Pyrimme seuraavassa valaisemaan asiaa luomalla kokonaiskuvaa erityyppisistä tutkimuksista käyttäen esimerkkejä. Haluamme korostaa *erilaisia* mahdollisuuksia huomioida hakujärjestelmän käyttäjä, kun tiedonhakua tutkitaan.

Yksittäisessä tutkimuksessa voidaan aina huomioida vain rajattu joukko hakijan piirteitä. Tiedonhaun tutkimusasetelmat voidaan karkeasti jakaa kolmeen tyyppiin, jotka kuvastavat eri tapoja suhtautua hakijaan. Suhtautumistavat voidaan otsikoida seuraavasti: (1) *autenttisten hakutilanteiden tutkiminen*, (2) *koehenkilöt suorittamassa heille annettuja tehtäviä* ja (3) *testikokoelmien hyödyntäminen*. Tyyppien 1 ja 2 tutkimuksissa tarvitaan tutkimukseen osallistuvia ihmisiä, joiden toimintaa tarkastellaan. Kolmanteen tyyppiin kuuluvia tutkimuksia voidaan sen sijaan tehdä ilman testihenkilöitä, hyödyntämällä *testikokoelmaa*.

## Autenttisten hakutilanteiden tutkiminen

Autenttinen tiedonhakutilanne tarkoittaa todellisen maailman tiedonhakua, jossa ihmiset tekevät tiedonhakuja luonnollisissa ympäristöissä, kuten työpaikoilla, liittyen oikeisiin tehtäviin, ongelmiin ja tilanteisiin. Tällainen tutkimus on harvinaista, sillä sen toteuttaminen on vaivalloista ja vaatii pääsyyntä erilaisiin tilanteisiin sekä runsaasti aikaa. Todellisen maailman asetelmilla voidaan kuitenkin löytää ja saada esille sellaista, mikä jäisi tutkimusasetelmien ulkopuolelle, jos asetelmat olisivat liian rajattuja. Tällaisia asioita voivat olla esimerkiksi tiedostamattomat tiedonhakutavat ja oikopolut, joista on tullut rutiineja työssä eikä niihin välttämättä kiinnitetä tietoisesti huomiota.<sup>3</sup> Tiedonhaun tutkimus lähestyy tällöin tiedonhankinnan tutkimusta ja tietokäytäntöjen tutkimusta, sillä huomioon otetaan tietojärjestelmien lisäksi myös muita tiedonlähteitä. Näitä voivat olla esimerkiksi painetut lähteet, omat muistiinpanot ja muut ihmiset, joilta kysytään neuvoja.

Usein autenttisia tiedonhakutilanteita tutkitaan työpaikoilla ja -tilanteissa. Informaatio näyttätty näissä varsin monimuotoisena. Työn muutos ja digitalisaatio ovat muuttaneet informaatioympäristöjä erilaisilla työpaikoilla. Hyvän informaatiotutkimuksen näkökul-

3 Kumpulainen 2017; Kumpulainen 2013.

masta tehdyn jäsenyyksen informaation roolista työpaikoilla tarjoaa Byströmin ja kumppanien teos<sup>4</sup>, jossa aiheen tutkimusta käsitellään monipuolisesti. Tutkimuksia on tehty eri aloilla, muun muassa sosiaalityön puitteissa,<sup>5</sup> kaupungin hallinnossa<sup>6</sup> sekä historian tutkijoiden työtehtävien parissa.<sup>7</sup> Kaikissa näissä eri alojen vaatimukset tietojärjestelmille ja niiden sisältämälle informaatiolle näyttäytyvät käyttötilanteiden kautta jokseenkin erilaisina. *Työn tavoitteet* määrittävät informaation hyödyllisyyttä. Tarkastelemme teemaa lisää luvussa 2.3, joka käsittelee informaatiovuorovaikutusta.

Autenttisten hakutilanteiden tutkimuksissa voidaan tarkastella myös vuorovaikutus- ja tiedonhakulokeja muun tutkimusaineiston (havainnointi, haastattelut ja lokit) tukena<sup>8</sup> tai tarkastella lokeja itessään,<sup>9</sup> jolloin tutkimuksen kohteena on usein suppeampi haku-tehtävä kuin edellä kuvatuissa työpaikoilla tehdyissä tutkimuksissa. Tällöin tarkastelutasona on usein työtehtävän kokonaisuus tai sen osa.

## Koehenkilöt suorittamassa annettuja tehtäviä

Autenttisten hakutilanteiden tutkimusta voidaan pitää tietyllä tapaa ihanteellisena. Parhaimmillaan asetelmasta puuttuu keinotekoisuus ja tutkija voi itse valita, millaisiin kysymyksiin hän keskittyy. Kaikkia tutkijaa kiinnostavia kysymyksiä ei kuitenkaan välttämättä voi tutkia autenttisissa tilanteissa. Sopivia tilanteita ei välttämättä ole tutkijalle saatavilla. Tällaisissa tilanteissa voidaan hyödyntää *simuloituja tehtäviä*.<sup>10</sup> Niissä koehenkilöt suorittavat autenttisten tehtävien sijasta *tutkijan antamia* tehtäviä. Tehtävänanto koostuu *taustatarinasta* sekä yksityiskohtaisemmista toimintaohjeista (kuvio 1).

4 Byström ym. 2019.

5 Huuskonen & Vakkari 2010.

6 Byström & Järvelin 1995; Saastamoinen & Kumpulainen 2014.

7 Korkeamäki & Kumpulainen 2019; Kumpulainen ym. 2020.

8 Ks. esim. Kumpulainen & Järvelin 2010; Saastamoinen & Järvelin 2017.

9 Odijk ym. 2015; White & Huang 2010.

10 Borlund 2016.

**Taustatarina**

- Tehtävä: kuvittele, että olet kirjoittamassa artikkelia, joka käsittelee Suomen historiaa tai maailmanhistoriaa 1800-luvun lopulla tai 1900-luvun alun näkökulmasta.
- Ohje relevanssiarvion tekemiseen on kuvattu alla.

**Hakutuloksen evaluointi**

- Relevanssitaso 3: Näkemäsi lehtileike sisältää runsaasti informaatiota hakemastasi aiheesta ("Bobrikoffin murha") ja lehtileike auttaa hyvin tehtävän suorittamisessa.
- Relevanssitaso 2: Lehtileike koskettaa hakemaasi aihetta ja auttaa jossain määrin tehtävän suorittamisessa.
- Relevanssitaso 1: Lehtileike liittyy aiheeseen, mutta se ei välttämättä auta suoraan tehtävän suorittamisessa.

**Kuvio 1.** Simuloidun tehtävän kuvaus (muokattu Kettusen ja muiden pohjalta).

Kuvio 1 havainnollistaa *simuloitua* hakutehtävää.<sup>11</sup> Koehenkilön tehtävänä on hakea hänelle annettua hakukonetta käyttäen digitoituja historiallisia sanomalehtileikkeitä ja *arvioida lehtileikkeiden hyödyllisyyttä taustatarinan valossa*. Kyseisessä tutkimuksessa paperille painetut historialliset sanomalehtiartikkelit näytettiin hakijalle digitaalisessa muodossa. Painettu teksti on kuitenkin jouduttu tunnistamaan optista merkintunnistusta käyttäen. Muodostettuun digitaaliseen tekstiin syntyy samalla tunnistusvirheitä. Tekstissä esiintyvät virheelliset kirjainmerkit haittaavat tiedonhakua sekä dokumenttien lukemista. Tutkimuksessa vertailtiin kahden eri merkintunnistumenetelmän (*optical character recognition*, OCR) vaikutuksia haakuun. Vertailussa olivat perusmenetelmä sekä vähemmän merkkivirheitä muodostava kehittyneempi versio. Koehenkilön käyttämä hakukone kykeni esittämään dokumentin kahtena vaihtoehtoisena versiona (joko perus- tai kehittyneempi OCR) ja keräämään talteen hakijan antaman palautteen dokumentin hyödyllisyydestä (hakija klikkasi dokumentille "laatuarvosanan" 0, 1, 2 tai 3). Tämä simuloitu

<sup>11</sup> Kettunen ym. 2022.

Hakuaiheet	Dokumenttikokoelma	Relevanssikorpus
Topic-id: #15 Aihe: Viiriäiset Suomessa Kuvaus: Relevantti (Rel=1) dokumentti sisältää tietoja viiriäistä koskevista havainnoista Suomessa ... Epärelevantti dokumentti (Rel=0) sisältää ... Topic-id ...	Docid: #300700 Otsikko: "Oletko nähnyt viiriäistä" Dokumentin teksti (5673): " ... kokkimme näki omin silmin viiriäisen käydessään ..." Docid: 300701 ...	Topic-id Docid Rel ... #15 #300601 0 #15 #300700 1 ... #16 #101315 0 ...

Kuvio 2. Testikokoelman komponentit: hakuaiheet, dokumentit ja relevanssiarviot.

asetelma teki mahdolliseksi tutkia, kuinka OCR-virheiden väheneminen vaikuttaa dokumenttien koettuun hyödyllisyyteen. Ilman simuloitua asetelmaa tällaista kysymystä olisi vaikea tutkia.<sup>12</sup>

## Testikokoelmatutkimus

*Testikokoelmatutkimus* havainnollistaa kolmatta tapaa suhtautua tiedonhakuun tutkimuksessa. Tätä jo 1960-luvulla niin sanotuissa Cranfield-kokeissa käytettyä mallia on hyödynnetty runsaasti tiedonhaun tutkimuksissa.<sup>13</sup> Testikokoelmatutkimuksessa koehenkilöitä ei välttämättä käytetä lainkaan. Tiedonhaun "testiajoja" voidaan suorittaa esimerkiksi omalla tietokoneella. *Testikokoelma* koostuu kolmesta rakenneosasta, jotka ovat (i) hakuaiheet (*topics*), (ii) dokumentit (*documents*) ja (iii) relevanssikorpus (*relevance corpus*). *Hakuaiheet* määrittelevät sen, mitä hakujärjestelmästä tarkkaan

<sup>12</sup> Kettunen ym. 2022.

<sup>13</sup> Cleverdon 1991; Voorhees 2007.

2 kyselyä (topic id #15)	Hakutulos 1 (q1)	Hakutulos 2 (q2)
<ul style="list-style-type: none"> <li>- q1: viiriäiset Suomessa</li> <li>- q2: viiriäiset</li> </ul>	<ul style="list-style-type: none"> <li>- 300700 Rel=1</li> <li>- 300601 Rel=0</li> </ul>	<ul style="list-style-type: none"> <li>- 300601 Rel=0</li> <li>- 300700 Rel=1</li> </ul>

**Kuvio 3.** Esimerkki testikokoelman hyödyntämisestä. Kahden kyselyn (q1 ja q2) tuottamia hakutuloksia voidaan vertailla hyödyntäen relevanssikorpuksen relevanssietietoja.

ottaen pyritään hakemaan. Hakuaiheita on tyypillisessä testikokoelmassa vähintään kymmeniä – tai jopa satoja. *Dokumentit* eli dokumenttitietokanta muodostavat hakujen kohteen. Tietokanta sisältää yleensä laajan joukon dokumentteja, ja vain hyvin pieni osuus dokumenteista on relevantteja jonkin yksittäisen hakuaiheen näkökulmasta. Kolmas rakenneosa, *relevanssikorpus*, sisältää relevanssiarviot eli tiedon siitä, mitkä dokumentit ovat tosiasiaa relevantteja kullekin hakuaiheelle (kuvio 2).

Kuviossa 2 on havainnollistettu testikokoelman komponentteja. Vasemmalla näkyvä *hakuaihe* (”Viiriäiset Suomessa”) määrittelee haun aiheen ja kuvaa, millainen dokumentin tulee olla, jotta se olisi relevantti. Kuviossa keskellä on havainnollistettu esimerkinomaisesti kokoelman yksittäistä dokumenttia. Hakuaihe ja dokumentti on merkitty yksilöivillä koodeilla (#15 ja #300700). Oikealla oleva *relevanssikorpus* kertoo, mitkä dokumentit ovat relevantteja kullekin aiheelle. Esimerkin ylin numerorivi kertoo, että *viiriäisaiheelle* (#15) dokumentti numero 300601 ei ole relevantti, sillä sen relevanssitaso on 0. Toinen rivi kertoo vastaavasti, että dokumentti numero 300700 on aiheelle relevantti, sillä sen relevanssitaso on 1. Testikokoelma rakennetaan käytännössä valitsemalla kohdedokumentit ja laatimalla sopivat hakuaiheet sekä relevanssiarviot. Relevanssiarviot muodostetaan niin sanotulla tuomarimenetelmällä, jolloin ihmiset lukevat dokumentteja ja arvioivat, ovatko ne relevantteja suhteessa valittuun aiheeseen. Testikokoelman valmistuttua sitä voidaan hyödyntää lukuisin tavoin tutkimuksessa. Yleisperiaatteena on muo-

dostaa kullekin hakuaiheelle kyselyitä ja hakea niiden avulla dokumentteja. Olennaista on oivaltaa, että hakutulosten *onnistuneisuutta* voidaan aina mitata *relevanssikorpuksen* ansiosta (kuvio 3).

Kuvio 3 havainnollistaa hakutulosten tarkastelua. Antoiko kysely q1 vai q2 paremman hakutuloksen? Kysely q1 palautti dokumentit #300700 ja #300601 tässä järjestyksessä ja (lyhyempi) kysely q2 palautti samat dokumentit, mutta päinvastaisessa järjestyksessä. Hakutuloksiin on kuviossa liitetty automaattisesti relevanssikorpuksen sisältämiä tietoja, jolloin tulosten paremmuutta voidaan verrata. Havaitsemme, että kyselyversio q1 palautti relevantin dokumentin ykkössijalla ja epärelevantin kakkossijalla. Intuitiivisesti arvioiden kysely q1 antoi hieman paremman tuloksen kuin q2. Tulosten täsmälliseen vertailuun on olemassa lukuisia numeerisia mittareita, jotka mallintavat esimerkiksi hakijan erilaisia preferenssejä. Yksi esimerkki tällaisesta mittarista on hakijan kärsivällisyys tutkia hakutulosta. Kuvio 3 havainnollistaa, kuinka testikokoelman avulla voitaisiin verrata esimerkiksi samasta aiheesta tehtyjen eri kyselyversioiden vaikutusta hakutulokseen.

Testikokoelmia hyödynnettäessä tehdään tyypillisesti tiedonhakuja kullakin (vähintään kahdella) menetelmällä, joita halutaan verrata, käyttäen useita hakuaiheita, pyrkimyksenä muodostaa yleisettäviä johtopäätöksiä. Tutkijaa kiinnostava kysymys voisi olla esimerkiksi, millaisia hakutuloksia lyhyet kyselyt tuottavat verrattuina pitkiin kyselyihin. Tällaista tutkimusasetelmaa on myös pelillistetty siten, että pelaajien syöttämät erilaiset kyselyt kilpailevat siitä, mitkä niistä tuottavat parhaita tuloksia.<sup>14</sup>

Kyselyjen ohella voidaan kokeellisesti tutkia myös esimerkiksi sitä, kuinka hyvin informaatiota *muodostavat* kielimallit, kuten ChatGPT, suoriutuvat inhimillisessä päätöksenteossa, vaikkapa ratkaistessaan oikeustapauksia.<sup>15</sup> Erilaisten kehoitteiden tai *kehoitestrategioiden* kuten ajatusketjujen *lopputulosta* päätöksentekoprosessissa

14 Arvola & Alamettälä 2022a; 2022b; Sormunen ym. 2002; Sormunen ym. 1998.

15 Esim. Yu ym. 2023.

voidaan yksinkertaisimmillaan evaluoida *tuomioistuinten tekemiä* langettavia tai vapauttavia *päätöksiä vasten*. Saatua tulosta analysoidaan määrällisesti eli kuinka usein spesifi kehotestrategia johtaa oikeaan ratkaisuun. Tämän kaltainen tutkimus on luonteeltaan *kysymys-vastaus*-tyyppisen (*Question Answering, QA*) tiedonhaun evaluointia, jossa hakukoneelle esitetään kysymys luonnollisella kielellä ja vastausten laatua arvioidaan verrattuna ihmisten laatimiin vastauksiin tai ratkaisuihin.<sup>16</sup>

Testikokoelmaa voidaan hyödyntää paitsi *kyselyiden ja kehotteiden* keskinäiseen vertailuun (tiedontarpeen ilmaisutapojen menestyksellisyys) myös *dokumenttien* ilmaisutapojen vertailuun (esimerkiksi kannattaisiko hakujärjestelmässä asettaa sanat haettaviksi taipuneissa muodoissaan tai perusmuotoisina) sekä *hakualgoritmien* vertailuun. Siinä, missä arjen tiedonhaussa tiedontarpeen ilmaisu, dokumenttien tietosisällön ilmaisu ja hakualgoritmi muodostavat erottamattoman kokonaisuuden, testikokoelman avulla voidaan tutkia kokonaisuuden yksittäisten osien vaikutusta haun tuloksiin. Perusidea on tällöin varioida yksittäistä asetelman komponenttia kerrallaan (vaikkapa *kyselyiden* tyyppiä) pitäen samalla asetelman muut pääkomponentit ennallaan (esimerkiksi indeksin tyyppi ja kyselyitä dokumentteihin vertaileva hakualgoritmi).

Testikokoelmaa hyödynnettäessä näkemys *hakijasta* pelkistyy voimakkaasti. Hakutilanteiden vertailu perustuu usein *yksittäisten* kyselyiden käyttöön hakuaiheille, jotka ovat täsmällisesti määriteltäviä, evaluoinnissa tarkastellaan haun lopputuloksia, relevanssi perustuu hakuaiheen ja dokumentin *aiheisältöjen* vastaavuuteen ja dokumentin relevanssi arvioidaan riippumatta muista samanaikaisesti haetuista dokumenteista. Todellisuudessa hakuprosessi on kuitenkin yleensä *olennaisesti vuorovaikutteinen*. Hakija voi turvautua useampiin kyselyihin, hakuaihe voi olla epäselvä ja hakijan relevanssikriteerit voivat olla subjektiivisia ja muuttuvia. Lisäksi hakija voi arvioida haettuja dokumentteja kokonaisuutena, ja muutkin seikat

16 Webber 2010; Voorhees & Tice 2000.

kuin hakutulosten relevanssi voivat olla olennaisia – esimerkiksi hakemisen vaivattomuus ja nopeus.

Testikokoelmia voidaan hyödyntää varsin monentyyppisen hakijoiden toiminnan tutkimiseen.<sup>17</sup> Esimerkkinä *monen kyselyn hakuistuntojen* simuloinnissa kuhunkin hakuaiheeseen liitetään yksittäisen kyselyn sijasta erilaisten kyselyiden sarja q1, q2, q3 ja niin edelleen. Tällöin testikokoelmaa käyttäen voidaan simuloida tilannetta, jossa hakija suorittaa hakuaiheelle useita kysely-yrityksiä seilaillen hakutulosta kunkin kyselyn jälkeen. Tällaisen asetelman erikoistapauksena voidaan pitää *relevanssipalautteen simulointia*, jolloin haetuista dokumenteista eristetään automaattisesti sanoja alkuperäiseen kyselyyn lisättäviksi. Testikokoelman käyttöalaa voidaan tähän tapaan pyrkiä laajentamaan, jolloin kokoelmaa hyödynnetään esimerkiksi sellaisen toiminnan jäljittelyyn, jossa hakija tarkastelee hakutulosta ja parantelee kyselyään lisäämällä siihen hyviä hakusanoja.

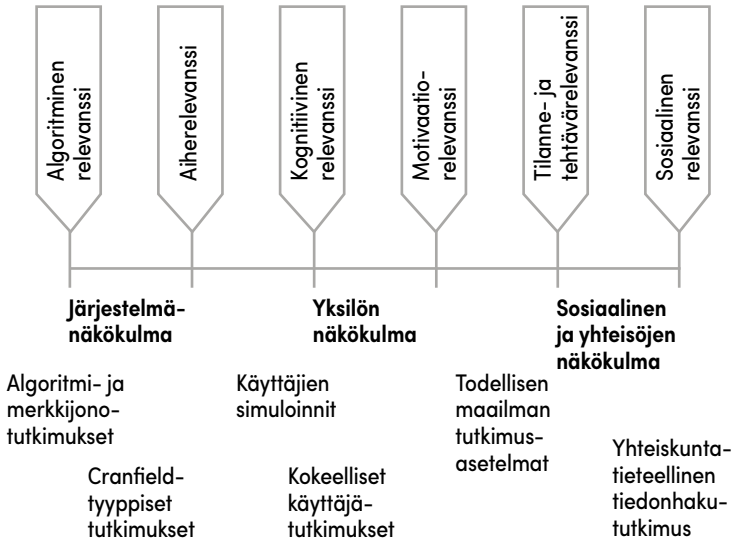
Testikokoelmien käytön etuna on, että niiden avulla voidaan systemaattisesti tarkastella hakujärjestelmän yksittäisten komponenttien vaikutusta haun tuloksiin yksinkertaistetussa asetelmassa. Mallin rajoituksia ovat hakuaiheiden, dokumenttikokoelman sekä relevanssiarvioiden rajallisuus ja malliin sisältyvät epärealistisen yksinkertaistavat oletukset, esimerkiksi relevanssin käsitteen redusoinnin informaation ”aiheenmukaisuudeksi”.<sup>18</sup>

## Relevanssi

Tiedonhaun tausta-ajatuksena on kulloiseenkin inhimilliseen tiedontarpeeseen sopivan tiedon tarjoaminen. Tiedonhaun tutkimus pyrkii yhä parempien tiedonhakujärjestelmien kehittämiseen. Sen määrittely, mikä on parempi järjestelmä, riippuu tarkastelukulmasta ja tutkimusasetelmasta. Tämä paremmuuden tarkastelu tarkoittaa

17 Ks. esim. Arvola ym. 2010; Baskaya ym. 2011; Keskustalo ym. 2009; Pääkkönen ym. 2017.

18 Lisätietoa mallin rajoitteista tarjoavat Turpin & Scholer 2006 ja Smith & Kantor 2008.



Kuvio 4. Relevanssin tyypit ja erilaisia tutkimusnäkökulmia.

tiedonhaun tuloksellisuuden eli onnistumisen tai epäonnistumisen tarkastelua. Tästä tarkastelusta puhutaan *tiedonhaun evaluointina*, jolloin jollain keinolla pyritään osoittamaan se, kuinka “hyvin” tiedonhaku toimii suhteessa tarkastelukulmaan ja tarkastelun kohteeseen. Tarkastelunäkökulmia voidaan rajata esimerkiksi aiemmin esitettyjen kolmen käyttäjäasetelman mukaisesti sekä tarkastelukohteen eli tiedonhaun osatekijöiden mukaan (kuvio 4).

Olellainen käsite tiedonhaun evaluoinnissa on *relevanssi*. Relevanssi tarkoittaa arviota informaation jonkin tyyppisestä “käyttökelpoisuudesta” jossain tilanteessa eli sitä, miten hyvin haettu tieto vastaa tiedontarpeeseen. Toisin sanoen relevanssi on suhde tiedontarpeen ja tiedon välillä.<sup>19</sup> Relevanssin määrittelyn tekee vaikeaksi se, että tiedontarpeen käsitteen määrittely vaihtelee eri tutkimuk-

19 Rieh & Xie 2006; Saracevic 1975.

sisä.<sup>20</sup> Relevanssi voidaan jakaa karkeasti *järjestelmien relevanssiin* ja *käyttäjärelevanssiin*.<sup>21</sup> Nämä esittävät ilmiöitä ikään kuin kahdessa eri maailmassa – *järjestelmien maailmassa* ja *käyttäjän (tiedon käyttäjän) maailmassa* – jotka ovat vuorovaikutuksessa keskenään. Puhutaankin *järjestelmälähtöisestä* ja *käyttäjälähtöisestä* tiedonhaun tutkimuksesta.

Kuviossa 4 on esitetty monipuolisempi jaottelu, jossa lähdetään algoritmista relevanssista kohti tiedon hakijaa ja hakijan toiminnan konteksteja. Voidaan ajatella, että siirryttäessä kuviossa vasemmalta oikealle tarkastelualue laajenee huomioimaan kasvavassa määrin inhimillisen kontekstin ulottuvuuksia. Relevanssista löytyy myös muita esityksiä ja jäsennyksiä.<sup>22</sup>

Kuvioon on merkitty myös tutkimustraditioita, joilla on erilainen näkökulma ilmiöihin. Jatkumon oikeassa ääripäässä tarkastellaan todellisen maailman ilmiöitä sellaisina kuin ne esiintyvät yhteisöissä, vasemmalla puolestaan on kuvattu kokeellinen ja voimakkaasti rajattu tiedonhaun laboratorioasetelma. Tällöin inhimillinen ulottuvuus ja ihmisen taustan huomiointi vähenee ja relevanssin määrittelmä vaihtuu ja liukuu toisenlaiseksi.<sup>23</sup>

Järjestelmä- eli *algoritmisen relevanssi* tarkoittaa tiedonhakupäätelmälle annetun *kyselyn* ja *dokumenttien esitysten* – käytännössä indeksien eli hakemistojen – välistä suhdetta eli sitä, miten merkijonoja täsmäytetään järjestelmässä ja onnistutaanko siinä (täsmäytyksestä, ks. luku 3.1) Jokaisella järjestelmällä on keinonsa esittää dokumentit, järjestää ne ja täsmäyttää eli verrata niitä kyselyyn. Tässä siis tarkastellaan sitä, miten järjestelmä päättää sen, liittyykö dokumentin informaatio annettuun kyselyyn.

*Aiherelevanssi* on varsin suosittu käsitteellistys tiedonhakupäätelmiä tarkasteltaessa, jolloin tiedonhakijan tilanteen ja taustan

20 Byström & Kumpulainen 2020.

21 Saracevic 2016.

22 Ks. esim. Belkin 2015; Borlund 2003; Cosijn & Ingwersen 2000; Ingwersen & Järvelin 2005; Saracevic 2007.

23 Kelly 2009.

tarkastelun osuus on kuvattu tutkimusasetelmassa vain hakutehtävän muodossa – joko simuloidun tilanteen kuvauksena<sup>24</sup> tai pelkistettynä hakuaiheeksi. Tämä tuo hallittavuutta ja toistettavuutta mittaamiseen, kun todellisen maailman monimuotoiset hakuaiheet ja hakutarpeet pelkistetään kontrolloiduiksi tehtäviksi, jotka toistuvat eri tutkimuksissa ja mittauksissa samoina. Aiherelevanssin mielessä haettu tieto (dokumentti) on relevanttia, jos se vastaa hakutehtävässä määriteltyä *aihetta*. Aiherelevanssin arviointi tapahtuu vertaamalla dokumentin informaatiosisältöä hakuaiheeseen tai hakutehtävään. Tällainen dokumentin tai informaation relevanssi on suhteellisen pysyvä ominaisuus. Esimerkiksi jos hakuaiheeksi on kuvattu ”Etsi tieto siitä, milloin Tampereen kaupunki on perustettu”, relevantti dokumentti kertoo vastauksen tähän – käyttötilannetta ei tarvitse ottaa huomioon relevanssin tarkastelussa. Aiherelevanssi on varsin suosittu käsitteellistys *hakujärjestelmiä* tarkasteltaessa. Sen etuna on, että tiedonhakijan tilanteen ja taustan tarkastelun osuus kuitistuvat tutkimusasetelmassa. Aiherelevanssi määritellään arvioimalla, *sisältääkö dokumentti informaatiota hakuaiheessa määritellystä aihepiiristä*. Aiherelevanssi arvioi vain dokumentin aiheenmukaisuutta, eikä sen arvioinnissa huomioida käyttäjien subjektiivisia tarpeita.

*Käyttäjärelevanssi* tarkoittaa tiedon hakijan *tiedontarpeisiin* vastaavaa informaatiota. Tässä voidaan ottaa huomioon joukko erilaisia piirteitä, jotka liittyvät esimerkiksi tiedonhakijan tietämyksen tilaan (*kognitiivinen relevanssi*), tunteisiin ja motivaatioon (*motivaatiorelevanssi*) tai hakutilanteeseen tai -tehtävään (*tilanne- ja tehtävärelevanssi*). *Hakijan tietämyksen tilan* piirteitä ovat aiempi, jo hankittu tieto sekä dokumentin tulkinta- ja tiedonhakutaidot. Tunteet ja motivaatio vaikuttavat myös siihen, kuinka informaatiota arvioidaan suhteessa tavoitteisiin ja millaista tietoa hyväksytään tai preferoidaan. Relevanssi on tiedon hakijan näkökulmasta katsottuna dynaamista, sillä tilanteesta ja tehtävästä opitaan tiedonhaun aikana.

24 Borlund 2003; 2016.

Samalla ymmärrys tiedontarpeesta muuttuu tiedon karttumisen myötä. Tiedontarpeiden katsotaan tilanne- ja tehtävärelevanssissa kumpuavan toiminnan kehyksestä eli tilanteen tai tehtävän kulloisestakin tulkinnasta. Lisäksi tilannesidonnaisuus määrittää tiedonhakua, jolloin esimerkiksi tilanteen ollessa ohi tiedontarve häviää ja haetun dokumentin relevanssi muuttuu. Tällainen tilanne voi olla esimerkiksi linja-auton aikataulun hakeminen – kun tarve on ohi, ei informaatiolla ole enää merkitystä tiedonhakijalle.

*Sosiaalisesta relevanssista* puhuttaessa tarkastellaan sitä, kuinka informaatio sekä informaatio- ja tietojärjestelmät vaikuttavat työhön, vapaa-aikaan, yhteiskuntaan ja kulttuuriin.<sup>25</sup> Sosiaalisen median alustat voidaan nähdä hakukoneina, ja niillä on ollut suuri vaikutus julkiseen keskusteluun ja yhteisöjen viestinnän muutokseen. Esimerkiksi COVID-19-epidemian aikana vuosina 2020–2021 liikkui runsaasti dis- ja misinformaatiota, joka ei ollut valtaväestön ja viranomaisten mielestä sopivaa tai informaation katsottiin olevan jopa haitallista yleiselle terveydelle. Tällä tiedolla oli vaikutusta erilaisiin yhteisöihin ja niiden yhteisöjen toimintaan, mutta se myös aiheutti paljon julkista keskustelua. Normatiivinen tieto tarkoittaa sitä, että tieto vastaa yhteiskunnassa vallitsevia normeja ja arvoja. Sosiaalisen relevanssin yhteydessä voidaan katsoa, että tieto on relevanttia, jos se on yleisesti hyväksyttyä ja hyväksytyjen auktoriteettien hyväksymää.

Tässä luvussa kuvasimme tiedonhakua ilmiönä, jonka ytimessä on pyrkimys yhdistää hakijan tiedontarpeen kuvauksia ja dokumenttien tietosisällön kuvauksia hyödyllisen tai relevantin informaation löytämiseksi. Korostimme artikkelissa näkemystä, että *todelliset* hakijat ja hakemisen käyttöyhteydet ovat moninaisia ja *tutkimuksen* aikana hakuprosessia ja hakijaa aina jotenkin *mallinnetaan*. Mallinnus sisältää valintoja ja on siksi lähtökohtaisesti rajaavaa. Mallinnus ei ole ongelma, josta tulisi päästä eroon, mutta valitun mallinnuksen rajoista olisi hyvä tulla tietoisiksi. Valaisimme asiaa

25 Saracevic 1995.

esittämällä jaon kolmentyyppisiin tiedonhaun tutkimuksiin. Ensimmäisessä tutkimustyyppissä tutkitaan ihmisten tiedonhakutoimintaa *autenttisissa* tilanteissa. Toinen tutkimuksen teon tapa hyödyntää koehenkilöitä, jotka suorittavat heille *annettuja* tehtäviä, jotka saattavat muistuttaa heille tuttuja tehtäviä tai olla enemmän tai vähemmän vieraita. Kolmanneksi tiedonhakua voidaan tutkia vielä edellisiä pelkistetympin testikokoelmaa käyttäen, jopa kokonaan *ilman* ihmisten läsnäoloa. Testikokoelman kolme peruskomponenttia ovat *hakuaiheet*, *dokumenttikokoelma* sekä *relevanssiarviot*. Kukin näistä kolmesta tutkimuksenteon tavasta ottaa omilla tavoillaan kantaa siihen, millaisena tutkimusasetelmaan sisältyvä tiedonhakija tutkimuksessa näyttäytyy eli minkä piirteiden kautta hakija tulee tutkimuksessa mallinnetuksi.

## Lähteet

- Arvola, Paavo & Tuulikki Alamettälä (2022a) IRVILAB: Gamified searching on multilingual Wikipedia. *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, 3329–3333.
- Arvola, Paavo & Tuulikki Alamettälä (2022b) Enhancing information literacy skills: A game design for seeking information and making queries. *Proceedings of games and learning alliance. GALA 2022. Lecture notes in computer science 13647*, Cham: Springer, 323–328.
- Arvola, Paavo & Jaana Kekäläinen & Marko Junkkari (2010) Expected reading effort in focused IR evaluation. *Information retrieval* 15:5, 460–484.
- Baskaya, Feza & Heikki Keskustalo & Kalervo Järvelin (2011) Simulating simple and fallible relevance feedback. *Proceedings of the 33<sup>rd</sup> European conference on advances in information retrieval*, 593–604.
- Belkin, Nicholas J. (2015) Salton award lecture: people, interacting with information. *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, 1–2. <https://doi.org/10.1145/2766462.2767854>
- Borlund, Pia (2003) The IIR evaluation model: A framework for evaluation of interactive information retrieval systems. *Information research* 8:3, 3–8.
- Borlund, Pia (2016) A study of the use of simulated work task situations in interactive information retrieval evaluations: A meta-evaluation. *Journal of documentation* 72:3, 394–413.
- Byström, Katriina & Kalervo Järvelin (1995) Task complexity affects information seeking and use. *Information processing & management* 31:2, 191–213.

- Byström, Katriina & Jannica Heinström & Ian Ruthven (toim.) (2019) *Information at work: Information management in the workplace*. Lontoo: Facet.
- Byström, Katriina & Sanna Kumpulainen (2020) Vertical and horizontal relationships amongst task-based information needs. *Information processing & management* 57:2, 102065. <https://doi.org/10.1016/J.IPM.2019.102065>
- Cleverdon, Cyril W. (1991) The significance of the Cranfield tests on index languages. *Proceedings of the 14<sup>th</sup> annual international ACM SIGIR conference on research and development in information retrieval*, 3–12.
- Cosijn, Erica & Peter Ingwersen (2000) Dimensions of relevance. *Information processing & management* 36:4, 533–550. [https://doi.org/10.1016/S0306-4573\(99\)00072-2](https://doi.org/10.1016/S0306-4573(99)00072-2)
- Huuskonen, Salla & Pertti Vakkari (2010) Client information system as an everyday information tool in child protection work. *Proceedings of the third symposium on information interaction in context*, 3–12.
- Ingwersen, Peter & Järvelin Kalervo (2005) *The Turn: Integration of information seeking and retrieval in context (The information retrieval series)*. Dordrecht: Springer.
- Korkeamäki, Laura & Sanna Kumpulainen (2019) Interacting with digital documents: A real life study of historians' task processes, actions and goals. *CHIIR 2019 - Proceedings of the 2019 conference on human information interaction and retrieval*. <https://doi.org/10.1145/3295750.3298931>
- Kelly, Diane (2009) Methods for evaluating interactive information retrieval systems with users. *Foundations and trends in information retrieval* 3:1–2, 1–224.
- Keskustalo, Heikki & Kalervo Järvelin & Ari Pirkola & Tarun Sharma & Marianne Lykke (2009) Test collection-based IR evaluation needs extension towards sessions – A case of extremely short queries. *Information retrieval technology: The 5<sup>th</sup> Asia information retrieval symposium (AIRS'09)*. LNCS 5839, 63–74.
- Kettunen, Kimmo & Heikki Keskustalo & Sanna Kumpulainen & Tuula Pääkkönen & Juha Rautiainen (2022) OCR quality affects perceived usefulness of historical newspaper clippings. A user study. Teoksessa Giorgio Maria di Nunzio & Beatrice Portelli & Dominic Redavid & Gianmaria Silvello (toim.) *Proceedings of the 18th Italian research conference on digital libraries (CEUR workshop proceedings 3160)*. CEUR-WS.org. <http://ceur-ws.org/Vol-3160/paper1.pdf>
- Kumpulainen, Sanna (2013) *Task-based information access in molecular medicine: Task performance, barriers, and searching within a heterogeneous information environment*. Väitöskirja. Tampere: Tampereen yliopisto. <http://urn.fi/URN:ISBN:978-951-44-9287-7>
- Kumpulainen, Sanna (2017) Task-based information searching. Teoksessa John D. McDonald & Michael Levine-Clark (toim.) *Encyclopedia of library and information sciences*. Boca Raton: CRC Press, 4526-4536. <https://doi.org/https://doi.org/10.1081/E-ELIS4>
- Kumpulainen, Sanna & Kalervo Järvelin (2010) Information interaction in molecular medicine: Integrated use of multiple channels. *IliX 2010* -

- Proceedings of the 2010 information interaction in context symposium*. <https://doi.org/10.1145/1840784.1840800>
- Kumpulainen, Sanna & Heikki Keskustalo & Boyang Zhang & Kostas Stefanidis (2020) Historical reasoning in authentic research tasks: Mapping cognitive and document spaces. *Journal of the association for information science and technology* 71:2, 230–241. <https://doi.org/10.1002/asi.24216>
- Odijk, Daan & Ryen White & Ahmed Hassan Awadallah & Susan Dumais (2015) Struggling and success in web search. *Proceedings of the 24th ACM international on conference on information and knowledge management*, 1551–1560.
- Pääkkönen, Teemu & Jaana Kekäläinen & Heikki Keskustalo & Leif Azzopardi & David Maxwell & Kalervo Järvelin (2017) Validating simulated interaction for retrieval evaluation. *Information retrieval journal* 20:4, 338–362.
- Rieh, Soo Young & Hong Iris Xie (2006) Analysis of multiple query reformulations on the web: The interactive information retrieval context. *Information processing & management* 42:3, 751–768.
- Saastamoinen, Miamaria & Kalervo Järvelin (2017) Search task features in work tasks of varying types and complexity. *Journal of the association for information science and technology* 68:5, 1111–1123. <https://doi.org/10.1002/asi.23766>
- Saastamoinen, Miamaria & Sanna Kumpulainen (2014) Expected and materialised information source use by municipal officials: Intertwining with task complexity. *Information research* 19:4. <http://InformationR.net/ir/19-4/paper646.html>
- Salton, Gerard & Michael J. McGill (1987) *Introduction to modern information retrieval*. Auckland: McGraw-Hill.
- Saracevic, Tefko (1975) Relevance: A review of and a framework for the thinking on the notion in information science. *Journal of the American society for information science* 26:6, 321–343.
- Saracevic, Tefko (1995) Evaluation of evaluation in information retrieval. *Proceedings of the 18th annual international ACM SIGIR conference on research and development in information retrieval*, 138–146. <https://doi.org/10.1145/215206.215351>
- Saracevic, Tefko (2007) Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American society for information science and technology* 51:13, 1915–1933. <https://doi.org/10.1002/asi.20682>
- Saracevic, Tefko (2016) The notion of relevance in information science: Everybody knows what relevance is. But, what is it really? *Synthesis lectures on information concepts, retrieval, and services* 8, i–109.
- Sormunen, Eero & Sakari Hokkanen & Petteri Kangaslampi & Petri Pyy & Bemmu Sepponen (2002) Query performance analyser - A web-based tool for IR research and instruction. *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval*, 450–450.
- Sormunen, Eero & Juha Laaksonen & Heikki Keskustalo & Jaana Kekäläinen & Harri Kempainen & Harri Laitinen & Ari Pirkola & Kalervo Järvelin (1998) IR-game - a tool for rapid query analysis in cross-language IR

- experiments. *Proceedings of the joint workshop on cross language issues in artificial intelligence & issues of cross cultural communication, 24th November, 1998, Kent Ridge Digital Labs, Singapore.* <https://jsaic.krdl.org.sg/pricai98>
- Smith, Catherine L. & Paul Kantor (2008) User adaptation: Good results from poor systems. *Proceedings of the 31<sup>st</sup> annual international ACM SIGIR conference on research and development in information retrieval*, 147–154.
- Turpin, Andrew & Falk Scholer (2006) User performance versus precision measures for simple search tasks. *Proceedings of the 29<sup>th</sup> annual international ACM SIGIR conference on research and development in information retrieval*, 11–18.
- Webber, Bonnie (2010) Question answering. Teoksessa Alexander Clark & Chris Fox & Shalom Lappin (toim.) *The handbook of computational linguistics and natural language processing*, 630–654. West Sussex: Blackwell Publishing Ltd.
- White, Ryan W. & Jeff Huang (2010) Assessing the scenic route: measuring the value of search trails in web logs. *Proceedings of the 33<sup>rd</sup> international ACM SIGIR conference on research and development in information retrieval*, 587–594. <https://doi.org/10.1145/1835449.1835548>
- Wolfram, Stephen (2023) *What is ChatGPT doing ... and why does it work?* Champaign: Wolfram Media.
- Voorhees, Ellen M. (2007) TREC: Continuing information retrieval's tradition of experimentation. *Communications of the ACM*. 50:11, 51–54.
- Voorhees, Ellen M. & Dawn M. Tice (2000) Building a question answering test collection. *Proceedings of the 23<sup>rd</sup> annual international ACM SIGIR conference on research and development in information retrieval*, 200–207. New York: Association for Computing Machinery. <https://doi.org/10.1145/345508.345577>
- Yu, Fangyi & Lee Quartey & Frank Schilder (2023) Exploring the effectiveness of prompt engineering for legal reasoning tasks. *Findings of the association for computational linguistics*: ACL 2023, 13582–13596. Toronto: Association for Computation.