

Katsaus tiedon tallennuksen ja haun tutkimuksen ongelmiin, suuntauksiin ja mahdollisuuksiin*

* Perustuu VTT:n informaatiopalvelulaitoksessa 13. 3. 1987 pidettyyn esitelmään *Tiedonhaun tutkimuksen merkityksestä käytännön tiedonhakutoiminnalle*.

Järvelin, Kalervo, Katsaus tiedon tallennuksen ja haun tutkimuksen ongelmiin, suuntauksiin ja mahdollisuuksiin: (A review on the problems, directions and potentials of the research on information retrieval). Kirjastotiede ja informatiikka 6 (2): 56—61, 1987.

The problems, directions and potentials of the research on information retrieval (IR) are reviewed from the viewpoint of practical information work. The types of research on IR are outlined on the basis of the IR process. The review focuses on the limitations of IR systems, empirical research on search skills and future possibilities provided by information technology. Although technology develops rapidly the same is not true of the conceptual understanding of the IR process. Knowledge concerning the end users is often not considered important in IR research. The review emphasizes, however, that this is invaluable from the practical viewpoint.

Address: University of Tampere, Department of Library and Information Science, P.O.Box 607, SF-33101 Tampere, Finland.

1. Johdanto

Tiedon tallennuksen ja haun tutkimus voidaan ymmärtää monella tavalla ja monessa eri laajuudessa. Otan tutkimuksen jäsentämisen lähtökohdaksi tiedonhauprosessin ja siihen kuuluvat osatekijät. Ainakin seuraavat osatekijät voidaan erottaa tiedonhauprosessissa: 1) tietojen, tietokantojen tuottajat, 2) tietokannat (olipa niiden tekninen toteutustapa ja sisältö mikä hyvänsä), 3) hakujärjestelmät (yleensä hakuvälineet) ja niiden kehittäjät ja myyjät, 4) käyttäjäliittymät, 5) järjestelmien (välittömät) käyttäjät, 6) loppukäyttäjät (tiedon tarvisijat) ja 7) hakujen tulokset ja niiden käyttö.

Tiedon tallennuksen ja haun koko prosessiin kuuluvat ainakin seuraavat vaiheet: 1) tietojen keruu, organisointi ja tallennus hakua varten, 2) hakupalvelujen luominen, kehittäminen ja markkinointi, 3) järjestelmien käyttäjien taitojen ja hakuvälineiden hankinta, 4) vuorovaikutus tietojen loppukäyttäjän kanssa, 5) haun valmistelu (strategian valinta), 6) haun suoritus, 7) tulosten arviointi ja toimitus loppukäyttäjälle ja 8) 'kirjanpito'.

Tarkoitukseni on tarkastella tiedon tallennuksen ja haun tutkimuksen ongelmia, suuntauksia ja mahdollisuuksia erityisesti *käytännön tiedonhaikutoiminnan näkökulmasta* — siis mitä käytännön vaikutuksia ja seurauksia on ongelmilla,

niiden ratkaisulla ja uusilla mahdollisuuksilla. Mikään yllä mainituista tiedonhakuprosessin osatekijöistä ei ole yksinkertainen tai muuntumaton, vaan kutakin löytyy monia eri tyyppisiä. Se, mitä tiedon tallennuksen ja haun tutkimusta pidetään käytännön tiedonhakutoiminnan kannalta merkittävänä tutkimuksena, riippuu siitä, mitä eri osatekijöiden ajatellaan olevan. Osatekijöiden tyyppi vaikuttaa mm. siihen, millaisiksi tiedonhakuprosessin vaiheet muotoutuvat ja mitä taitoja tiedon hakijalta vaaditaan.

Käytännön tiedonhakutoimintaan kuuluviksi katson tiedonhakutoiminnan *johtamisen* (suunnittelu, kehittäminen, päätöksenteko, seuranta), tiedonhakutoiminnan *organisoinnin* (toimeenpano, välineiden, rahoituksen, henkilöiden ja taitojen hankinta) ja tiedonhakujen *suorittamisen* (hakujen valmistelu, suorittaminen ja tulosten jakelu).

Tämän jäsentelyn jälkeen voidaan kysyä, millaisella tiedon tallennuksen ja haun piirissä tehdyllä tutkimuksella on merkitystä käytännön tiedonhakutoiminnan kannalta. Vastaus on, tietysti: *kaikella hyvällä tiedonhaun tutkimuksella, jos ei välittömästi, niin ainakin ennenpitkää*. En tunne sellaista tiedonhaun tutkimusalueita, josta uskaltaisni sanoa, että tuolla nyt ainakaan ei ole, eikä tule olemaan, merkitystä käytännön tiedonhakutoiminnalle. Mutta, jos pyritään tekemään (tai epäsuorasti aikaansaamaan) onnistuneita tiedonhakuja, ei käytännön kannalta kaikki merkittävä tutkimus suinkaan ole suoranaisesti ja vain tiedon tallennuksen ja haun tutkimusta. Jokaisen palvelun onnistumisen edellytys on *keskitettyminen asiakkaisiin*. Näin informaatiopalvelusakin. Lisäksi jokainen hyvä kauppias *tuntee perusteellisesti myymänsä tuotteet*: tämän takia tiedonhakutoiminnan onnistuminen riippuu myös dokumenttien tai muiden data/informaatiokokonaisuuksien tuntemisesta (ominaisuudet, tuotantotavat). Jäsenen seuraavaksi tutkimusalueita tiedonhakuprosessin jäsenyyksen ja näiden lisänäkökohtien avulla ja esitän, missä mielessä kukin tutkimusalue on tarpeellinen käytännön tiedonhakutoiminnassa.

- *Volyymitutkimukset*: toiminnan ja sen osaluokkien määrällinen kehitys. Näitä tarvitaan toiminnan mitoittamista varten.

- *Taloustutkimukset*: toiminnan ja sen osaluokkien kustannukset tai hinnat siihen osallistuvien tahojen kannalta; ennusteet kustannusten kehityksestä. Näitä tarvitaan toiminnan rahoitusta ja markkinointia varten.

- *Asiakas- tai käyttäjätutkimukset*: ketkä ovat tiedonhakujen loppukäyttäjät ja välittömiä käyttäjiä nyt ja tulevaisuudessa; millaisia he ovat ja miksi he tiedonhakuja tarvitsevat; entä ketkä ei-

vät käytä ja miksi; mitä hyötyä heillä on tiedonhauista; miten ne vaikuttavat. Näitä tarvitaan asiakkaiden toiminnan ymmärtämiseksi ja palvelumuotojen kehittämiseksi.

- *Liittymä- ja välinetutkimukset*: mitä tekniisiä, menetelmällisiä ja käsitteellisiä välineitä on tai tulisi kehittää ensinnäkin tietojen *esittämiseen*, toiseksi hakujen *muotoiluun* ja kolmanneksi haun aikana saatavaan *opastukseen* nykyisin ja tulevaisuudessa. Työvälineet on tunnettava. Välineiden ominaisuuksista riippuvat niiden käyttäjiltä vaadittavat taidot ja se, ketkä ylipäätään voivat tulla käyttäjiksi.

- *Tiedon sisältö- ja organisointitutkimukset*: mitä tietokantoihin on pantu ja mitä tullaan lähivuosina panemaan; miten tiedot organisoidaan nyt ja tulevaisuudessa. »Tuotteet» on tunnettava. Ammattimaisten tiedonhakijoiden on tunnettava käsitteellinen organisointitapa, jotta he kykenevät muotoilemaan hakuja. Lisäksi ammattimaisten tiedonhakijoiden on ymmärrettävä tietojen tekninen organisointitapa niin pitkälle, että he käsittävät yleisellä tasolla haun vaatiman käsitteeltävän — sehän taas vaikuttaa mm. haun nopeuteen ja hintaan. Kaikki tietokannat eivät tule perustumaan käänteistiedoston käyttöön. Kaikki tietojen väliset suhteet eivät tule perustumaan hierarkisiin luokkasuhteisiin tai tesaurussten tarjoamiin suhteisiin eikä kaikki dokumenttien kuvailu tule perustumaan avainsanoihin tai luokitusysoleihin.

- *Hakumenetelmätutkimukset*: tiedon organisointi ja haku kulkevat käsi kädessä. Tämän alueen merkitys ilmenee edellisestä kohdasta. Sekä tekniset että käsitteelliset hakutavat ja niiden kehitys on hallittava.

- *Perustekniikkoja* koskevat tutkimukset ja kehitystyö: muistitekniikka, viestintätekniikka, käsitystekniikka ja esitystekniikka ja näiden kehitysnäkymät. Tiedonhaun kehitys on oleellisesti riippunut ja tulee edelleen riippumaan perustekniikoiden tarjoamista mahdollisuuksista.

Kaikista näistä kysymyksistä on tarjolla tutkimusta, toisista teemoista enemmän, toisista vähemmän. Kaikki tällainen tutkimus on jostain näkökulmasta merkittävää käytännön tiedonhakutoiminnan kannalta. Tutkimuksen merkitys käytännön kannalta voidaan vielä jäsentää seuraaviin kahteen ryhmään: 1) tutkimus, joka ratkaisee nykyisiä ongelmia, ja 2) tutkimus, joka luovuttaa mahdollisuuksia — muuttaa toimintaa.

Tarkastele seuraavaksi joitakin nykyisen tiedonhakutoiminnan ongelmia. Sen jälkeen syvennän neljän viimeksimainitun tutkimusalueen tarkastelua.

2. Nykyisten tiedonhakujärjestelmien ongelmia

Palvelussa asiakaskeksisyys on oleellista. Siksi jätän ongelmien tarkastelussa tiedonhakuutoiminnan sisäiset ongelmat (esim. suunnittelu, organisointi) huomiotta ja keskityn hakujen suorittamiseen. Loppukäyttäjän kannalta perinteiset tiedonhakujärjestelmät ovat monella tavalla rajoittuneita:

- *niiden käyttö on vaivalloista*, koska se
 - vaatii erikoislaitteita, joita on toistaiseksi harvassa,
 - vaatii erikoistietoja tai henkilöitä, joita on harvassa,
 - vaatii tiedontarpeen muokkaamista kysymykseksi.
- *niiden tarjoamat tulokset ovat vaatimattomia*:
 - ne tarjoavat käyttäjilleen kirjallisuusviitteitä, eivätkä varsinaista tietoa, jota loppukäyttäjät viimekädessä tarvitsevat;
 - viitteiden jälkeen on vielä suuri vaiva viitatujen dokumenttien hankinnassa ja vastaus-ten etsimisessä ja lukemisessa niistä; viitteiden perusteella ei voida varmasti ratkaista sitä, mitkä dokumentit ovat relevantteja (mistä vastaus löytyy, jos ollenkaan löytyy);
 - pelkässä viitteenhaussakin tulokset ovat vaatimattomia: pitkien viitelistojen hyödyttömyys on havaittu (hankittavaksi poimitaan tavallisesti vain muutama viite pitkän listan alusta); saanti ja tarkkuus eivät ole 100 % ja 100 %; yksimielisyyttä puuttuu mm. te-saurusten laatijoiden, indeksoijien ja hakijoiden kesken sopivista termeistä sekä ja loppukäyttäjien kesken dokumenttien relevanssista⁴;
 - kaikkia tarpeellisia tietotyyppejä ei edes löydy viitteiden kuvaamista dokumenteista²¹.

Tiedon hakija kohtaa ainakin kolme ongelmaa. Ensinnäkin luonnollisella kielellä esitetty kysely on epätäydellinen, epätäsmällinen ja moniselitteinen. Toiseksi sopivia, mahdollisia hakusanoja on lukuisia. Kolmanneksi kyselyohjelmien täsmäytymekanismi on rajoittunut: tavallisessa Boolean logiikkaan perustuvassa haussa se on dikotominen (täsmää/ei täsmää), minkä takia *melkein täsmääviä* dokumentteja ei löydetä. Näistä syistä hakijalle jää vastuu kysymyksen ymmärtämisestä, oikeiden termien huomaamisesta ja sopivan hakustrategian luomisesta hakuohjelman ja tietokannan antamien rajoitusten puitteista. Lyhyesti sanottuna tiedon tallennus- ja hakujärjestelmät ovat tyhmiä järjestelmiä fiksuille käyttäjille.⁷

Tiedon tallennuksen ja haun tutkimuksen puolella näitä ongelmia vastaa tiedon tallennuksen ja haun perusdilemma: Yhtäältä ei ole olemassa riittävää teoriaa kielestä ja merkityksestä luonnollisella kielellä esitetyn tekstin (dokumenttien) käsittelyyn. Toisaalta samasta asiasta voidaan puhua ja puhutaan lukemattomilla eri tavoilla (kirjoittajat, abstrahoijat, indeksoijat, hakijat, asiakkaat). Lisäksi täsmäytymekanismi on edellä kuvattu ongelma ja eri toimijoiden subjektiiviset relevanssiarviot poikkeavat melkoisesti toisistaan.⁷

Nämä ongelmat ja rajoitteet vähenevät ajan mittaan tiedonhakutekniikan, välineiden tiedon tallennuksen ja haun prosessin käsitteellisen hallinnan ja tietokantojen kehittymisen mukana. Joitakin ongelmia voidaan voittaa tuottamalla tietokantoja, joista käyttäjä saa viitteiden sijasta suoraan tarvitsemansa tiedot käytettäväkseen. Näitä tietokantoja on monentyyppisiä (mm. *tekstitietokannat*²⁵ ja *numeeriset tietokannat*^{3, 14, 17 20}).

Niiden määrät kasvavat nopeasti ja ne ovatkin nykyisin viitetietokantoja tärkeämpiä, mikäli esim. liikevaihtoa pidetään tärkeyden mittana. Lukumäärien mukaan tekstitietokantoja, numeerisia tietokantoja ja kirjallisuusviitetietokantoja on kutakin runsas neljännes kaikista, runsaasta 3200 julkisesta onlinetietokannasta, loppujen ollessa mm. ohjelmisto- sekä projekti- ja organisaatioviitetietokantoja^{5, 6, 10}. Otaksun, että lähivuosina ponnistellaan kovasti tietokantojen integroinnin kimpussa: pyritään saamaan erityyppiset tietokannat toimiviksi kokonaisuusiksi.

Joitakin ongelmista voidaan ratkoa kehittämällä tiedon tallennuksen ja haun välineitä ja joitakin taas käyttäjäliittymiä (interfaces) kehittämällä.

3. Hakutaitoon liittyvä tutkimus

Hakutaitoon liittyvä tutkimus suuntautuu kahtaalle: toisaalta järjestelmien käyttäjien (tiedonhakijoiden) käyttäytymisen, taidon ja virheiden empiiriseen tutkimiseen sinänsä ja toisaalta taidon siirtoon tiedonhaun välittäjäjärjestelmiin — siis tiedonhaun liittymien kehittämiseen.

Olemassa olevasta empiirisestä tutkimuksesta voi jokainen tiedonhakija ottaa oppia. Tutkimustuloksissa on tarjolla tietoa esim.^{1, 11, 12, 16}:

- tiedonhakijoiden yleisimmin käyttämistä hakustrategioista,
- hakijoiden käytettävissä olevista hakujen muuntelukeinoista ja niiden tavallisista käyttötavoista,
- hakijoiden tavallisista virheistä ja niiden syistä.

Käyttäjälittyneiden tutkimuksessa on nykyisin keskeistä automaattisten, usein tekoälyyn perustuvien välineiden kehittäminen helpottamaan tiedonhakuprosessin eri vaiheita. Suurin ongelma koskee olemassa olevan teknologian soveltamista tiedonhakuun — uusia laitteita ei niinkään tarvita, vaan pikemminkin jo olemassa olevan laitteiston ja ohjelmistotekniikan käyttötapojen määrittämistä. Ongelmien painopiste on siirtynyt toisaalta tiedonhakuprosessien käyttäjien ja toisaalta luonnollisen kielen käsittelyn puolelle. Kehitys riippuu siitä, kuinka hyvin ymmärretään toisaalta eritasoisten tiedonhakijoiden taitoja ja jäsennyksiä (mentaalimalleja) tiedonhausta, hakuohjelmista ja tietokannoista ja toisaalta luonnollista kieltä.^{2, 7, 8, 13}

Nykyiset tiedonhaun apuneuvot tukevat pääasiassa tiedonhaun matalan tason ongelmien ratkaisua. Tällä tarkoitetaan tehtäviä kuten yhteyden automaattinen luominen, komentokielten piilottamien käyttäjältä, kommentojen kirjoittamisen yksinkertaistaminen (esim. termikaaviot ja termin katkaisu), automaattinen peruutus virhetilanteista, yksinkertaisen standardihaun automaattinen suorittaminen, jne. Erilaisten *työkälujen* tarjoaminen ei yksin riitä: käyttäjillä tulee olla myös tieto siitä, milloin, mihin tarkoitukseen, miten ja millä seurauksilla työkaluja käytetään.

Tiedonhaun tuloksellisuuden nostaminen edellyttää sellaisten välittäjäjärjestelmien luontia, jotka avustavat haun aiheen analysoinnissa, tietokantojen valinnassa ja hakustrategian suunnittelussa²². Pitemmälle päästään, jos järjestelmiin voidaan sisällyttää tietoa ja taitoa mallittaa eri tyyppisiä käyttäjiä ja sopeutua heidän eroihinsa, ottaa haun aihepiiriin käsitteellinen rakenne huomioon sekä sovittaa haku kunkin tietokannan loogisen rakenteen mukaisesti^{23, 24}. Kun nykyiset hakuohjelmat pääasiassa tarjoavat tyhmiä välineitä fiksuille käyttäjille, voidaan tekoälyn keinoin nostaa joko hakuohjelmien tai välittäjäjärjestelmien (tai molempien) fiksuutta. Online tiedonhakua taitamattoman loppukäyttäjän mahdollisuudet hyvien hakutulosten saamiseen paranevat.

4. Teknisiä tulevaisuudennäkymiä

Tietotekniikan kehitys muokkaa tietokannoista uudentyyppejä informaatiojärjestelmiä, jotka tulevat voimakkaasti muuttamaan »tietoteollisuutta» ja informaatio businessia, kun tiedon käyttäjät, tuottajat ja julkaisijat mukautuvat uuteen tilanteeseen. Uudentyyppisten informaatiojärjestelmien kehitys perustuu suurelta osalta

kolmen perustekniikan kehitykseen: mikrotietokoneet, muistivälineet ja tietoliikennetekniikka. Selostan seuraavassa pääasiassa Foxin¹⁵ käsityksiä perustekniikoiden kehittymisestä ja niiden vaikutuksista. Aluksi havaintoja ja ennusteita eri perustekniikoiden kehityksestä.

Mikrotietokoneiden kehitys

Mikrotietokoneiden ja työasemien välinen kulu katoaa nopeasti. Nykyiset kehittyneet mikrotarjoajat käyttäjilleen jo pienen työaseman ominaisuudet:

- keskusmuistia 1 Mb (megatavua — vastaa noin 500 tavallista tekstisivua; valtaosan varaa tietokoneen ohjelmisto)
- laskentanopeus 1 MIPS (miljoona käskyä sekunnissa)
- melko nopeat levymuistit, joiden kapasiteetti on 10—40Mb (siis jopa 20 000 tekstisivua)
- melko erottelykykyinen grafiikka (640 × 480 pistettä vs. työasemien noin 1000 × 800 pistettä)
- ikkunointi, moniajo
- kehittyneet osoittamisvälineet (kuten hiiri)
- *sovellukset*: kuvan, tekstin, tiedonhallinnan ja taulukkolaskennan integrointi
- tulossa ovat tehokkaat ja monipuoliset tekoälyyn perustuvat välineet päätöksenteon ja tiedonhaun tueksi.

Muistivälineiden kehitys

Vaikka magneettilevyillä onkin suuria tietokantoja (aina 1 Gb asti; Gb = gigatavu), on niiden taloudellinen tallennuskapasiteetti rajallinen (ja rajoillaan). Optiset muistivälineet tuovat tähän ratkaisun:

- suurten optisten muistiyksikköjen kapasiteetti on jopa 128 Gb ja koko vain pienen auton koko; 128 Gb vastaa 320 000 kirjaa, joissa kussakin on 200 A4-sivua
- Pienempien CD-levyjen (5 tuumaa) kapasiteetti on 0.55 Gb, mikä vastaa yli 1000 kirjan koko tekstin tallennuskapasitettia;
- vaikka toistaiseksi optista muistia ei voida käyttää uudelleen (se on *read-only* tai *non-erasable*), poistuneen tämän rajoituksen muuttaman vuoden kuluessa.

Tietoliikennetekniikan kehitys

Pitkän matkan pakettiverkot (packet switching networks) ja lähiverkot (LANs, local area

networks) tarjoavat jo nykyisinkin virheettömän tekstin ja datan siirtomahdollisuuden nopeuksilla 2400. . .9600 bps. Lähivuosina aletaan toteuttaa ISDN-verkkoa (International Standard Data Net), joka tulee tarjoamaan tekstin, datan, kuvan ja puheen integroidun viestintämahdollisuuden aina 64 Kbps nopeuksiin asti. Tällä nopeudella voidaan teksti ja liikkumaton kuva lähettää nopeasti, mutta esim. tavallisen TV- uutislahetyksen liikkuvan kuvan lähettäminen veisi moninkertaisesti uutisten tavallisen kestoajan.

Kehityslinjoiden merkitys

Mikrojen kehitys aiheuttanee seuraavia vaikutuksia:

- datan ja sen käsittelyn hajautus lisääntyy;
- tietokantojen tuotanto ja jakelu hajautuu;
- downloading ja uploading yleistyvät;
- paikalliset hakuekspertit (asiantuntija- tai totokijärjestelmät) yleistyvät: tietokantojen valinta, yhteydenotto, haun muotoilu, käsittely ja esitys kuuluvat niiden toimintoihin — tiedonhaku saadaan riittävän helpoksi loppukäyttäjää varten.

Muistivälineiden kehityksestä seurannee mm.:

- full-text tietokannat tulevat laajamuotoisina taloudellisesti mahdollisiksi;
- kuva-aineisto, jonka tilantarve on noin 10-kertainen vastaavan kokoiseen tekstiin verrattuna, saadaan mukaan tietokantoihin;
- isot tietokannat (koko kuitenkin alle 1 Gb) jaetaan yleisesti optisilla levykkeillä suurkäyttäjille paikallista käyttöä varten;
- monet suuret hakuteokset ovat käytettävissä optisilla levyillä.

Tietoliikennetekniikan kehitys aiheuttanee mm.:

- full-text tietokantojen sisältämiä dokumentteja kuvineen voidaan välittää yleis- ja paikallisverkoissa nopeasti ja virheettää;
- verkkojen kytkentä paikallistasolta maailman tasolle tulee vielä nykyistä helpommaksi — läpinäkyväksi (transparent).

Näiden seurausten vaikutuksesta mm.:

- keskitettyjen hakujärjestelmien käyttäjämäärät kasvavat, mutta käytetty yhteisaika ei kuitenkaan kasva vastaavasti; tämän takia tietokantojen käytön laskutuspolitiikat muuttuvat täydellisesti; yhteysajan käytöstä laskutusperusteena luovutaan (ks. myös^{9, 18, 19, 20});
- downloading aiheuttaa lisääntyviä copyright-ongelmia ja varkauksia; laki ja tietojen hinnoittelu muuttuvat myös tämän takia.

Näyttää siltä, että tekniikan osalta käytännön tiedonhaun kehittämismahdollisuudet paranevat nopeasti. Tekniset edistysaskeleet tällä alueella eivät kuitenkaan sinänsä ratkaise kuin pienen osan tiedonhaun teoreettisista ja käytännön ongelmista: edellytykset teknisten mahdollisuuksien hyödyntämiseen eivät ole kasvaneet yhtä huomasti! Tietoliikenne- ja muun tietotekniikan edistysaskeleet eivät ratkaise tiedon tallennuksen ja haun ongelmista muita kuin sellaiset, joissa on vaivannut teknisen kapasiteetin tai taloudellisen toteuttamismahdollisuuden puute. Jäljelle jää suurin osa ongelmista ja lisäksi syntyy uusia. Näihin kuuluvat tiedonhaun käsitteellisen hallinnan ongelmat (tiedon organisointi ja haku) ja sosiaaliset ongelmat. Tiedon tallennuksen ja haun tutkimuksessa tarvitaan suuria ponnistuksia teknisten mahdollisuuksien täysimittaiseksi hyödyntämiseksi.

5. Yhteenveto

Olen jäsentänyt tiedonhaun tutkimusta tarkoituksenani tarkastella sen merkitystä käytännön tiedonhakutoiminnan kannalta. Tätä varten jäsentäsin käytännönkin johtamiseen, toimeenpanoon ja suoritukseen. Merkitystä koskeva tarkasteluni oli osittain kuin johtajan ns. nahkapäätös: kaikki hyvä tutkimus on merkittävää, jollei välittömästi, niin ainakin pitkän ajan kuluessa. Ajalliset muutokset (kehityslinjat) ja tämänhetkiset ongelmat (ja niiden ratkaisuyritykset) ovat keskeisiä. Viisas varautuu muutoksiin ennakolta ja oppii virheistään.

Jos halutaan nimenomaan pyrkiä onnistuneisiin tiedonhakuihin, on kiinnitettävä huomiota seuraaviin seikkoihin: *Ensinnäkin* tulee selkeästi tiedostaa tiedonhaun käsitteellisen hallinnan kehityksen hitaus verrattuna teknisten mahdollisuuksien kehitykseen. Merkitystä on erityisesti tutkimuksella, joka kuvaa, miten ja mihin huikeita keinoja voitaisiin käyttää. Odotan edistystä sekä tiedon organisoinnissa että älykkäässä haussa ja päättelyssä. *Toiseksi* tulee kiinnittää huomiota hakutaitojen empiiriseen tutkimukseen. Sitä on tehty ja tullaan tekemään lisää. *Kolmanneksi* on muistettava, että peli on menetetty, jos välittäjä, olipa kyseessä koneellinen tai inhimillinen välittäjä, ei ymmärrä loppukäyttäjän ongelmia ja tarpeita tai ei pysty sopeutumaan hänen kommunikointitapaansa. Ehkä tärkein tutkimus onnistuneiden hakujen kannalta on se, joka lisää välittäjän tietämystä loppukäyttäjän tehtävistä, ammattitaidosta, työtaivoista ja -menetelmistä. Tätä ei yleensä lasketa varsinaiseksi tiedonhaun tutkimukseksi. Käytännössä tällaisen tiedon hankinta ratkaistaan usein valitsemalla in-

formaatikoksi sopivan koulutustaustan omaava henkilö, joka siis *ymmärtää* loppukäyttäjää. Tiedonhaun tutkimusta loppukäyttäjien tutkimuksesta siitä tulee silloin, kun koulun penkille pannaan välittäjäjärjestelmä — asiantuntijajärjestelmä, jolle loppukäyttäjien tuntemusta ja ymmärtämiskykyä yritetään opettaa.

Hyväksytyt julkaistavaksi 6. 5. 1987.

Viitteet

¹ Bellardo, Trudi »What do we really know about online searchers?», *Online Review*. 9(3): 223—239; 1985.

² Brooks, H. M.; Daniels, P. J.; Belkin, N. J. »Research on information interaction and intelligent information provision mechanisms», *Journal of Information Science* 12: 13—44; 1986.

³ Chen, C.-c.; Herson, P. eds. *Numeric Databases*. Norwood, NJ: Ablex, 1984.

⁴ Cleverdon, Cyril »Optimizing convenient online access to bibliographic databases», *Information Services & Use* 4: 37—47; 1984.

⁵ *Directory of Online Databases*. Santa Monica, CA: Cuadra Associates; 1984.

⁶ *Directory of Online Databases*. Santa Monica, CA: Cuadra Associates; 1987.

⁷ Doszkoes, Tamas E. »Natural Language Processing in Information Retrieval», *Journal of the ASIS*. 37(4): 191—196; 1986.

⁸ Doszkoes, Tamas E. »IR, NLP, AI and UFOs: or IR-relevance, Natural Language Problems, Artful Intelligence and User-Friendly Online Systems» In: Rabitti, Fausto ed. *Proceedings of the 1986 ACM Conference on Research and Development in Information Retrieval*, Pisa, Italy, Sept. 8—10, 1986; 1986: 49—57.

⁹ Dunn, R. G.; Boyle, H. F. »Online Searching: an Analysis of Marketing Issues», *Information Services & Use*. 4: 147—154; 1984.

¹⁰ *EUSIDIC Database Guide 1983*. Learned Information, Oxford, 1983.

¹¹ Fidel, Raya »Online Searching Styles: A Case-Study-

Based Model of Searching Behavior», *Journal of the ASIS*. 35(4): 211—221; 1984.

¹² Fidel, Raya »Moves in online searching», *Online Review*. 9(1): 61—74; 1985.

¹³ Fidel, Raya »Towards Expert Systems for the Selection of Search Keys», *Journal of the ASIS*. 37(1): 37—44; 1986.

¹⁴ Foster, A. »Business Information from Databanks: the Potential of Online Numeric Databases», *Business Information Review*. 1(1): 38—45; 1984.

¹⁵ Fox, Cristopher »Information System Evolution in the Near Future», *Journal of the ASIS*. 37(4): 215—220; 1986.

¹⁶ Gaul, Kathleen »Learning to search: how people become database searchers» *Online Review*. 10(6): 355—367; 1986.

¹⁷ Heim, K. M. ed. *Data Libraries for the Social Sciences. A Special Issue of Library Trends*. 30(3); 1982.

¹⁸ Hull, D. »Marketing and Pricing of Full-text End-user Services», *Information Services & Use*. 4: 167—170; 1984.

¹⁹ Hunter, J. A. »What Price Information», *Information Services & Use*. 4: 217—223; 1984.

²⁰ Järvelin, Kalervo User charge estimation in numeric online databases: a methodology. *Acta Universitatis Tampereensis*, ser A vol 212. Tampere, Finland: University of Tampere; 1986.

²¹ Järvelin, Kalervo »Kaksi yksinkertaista jäsenystä tiedon hankinnan tutkimiseen», *Kirjastotiede ja informatiikka*. 6(1); 1987 (in press).

²² Marcus, R. S. »An Experimental Comparison of the Effectiveness of Computers and Humans as Search Intermediaries», *Journal of the ASIS*. 34(6): 381—404; 1983.

²³ Pollit, A. S. »Expert systems and the information intermediary: Tackling some of the problems of naive end-user search specification and formulation» In: Brookes, B. C. ed. *Intelligent Information Systems for the Information Society: Proceedings of the IRFIS 6 Conference*, Frascati, Italy, Sept. 16—18, 1985. Amsterdam, NL: North-Holland; 1986: 100—108.

²⁴ Salton, Gerard »Recent Trends in Automatic Information Retrieval». In: Rabitti, Fausto ed. *Proceedings of the 1986 ACM Conference on Research and Development in Information Retrieval*, Pisa, Italy, Sept. 8—10, 1986; 1986: 1—10.

²⁵ Tenopir, C. »Full-Text Databases» In: Williams, M. E. ed. *Annual Review of Information Science and Technology*, vol. 19. White Plains, NY: Knowledge Industry Publications; 1984: 215—246.