

KALERVO JÄRVELIN

Katsaus elektronisen tiedonvälityksen teknologiaan

Artikkelissa kuvataan tietokantatoimittamista ja -julkaisemista uusina joukkotiedotuksen dokumenttien laatimisen ja jakelun muotoina useiden tyyppisten vastaanottimien avulla tapahtuvaa käyttöä varten. Tarkemmin eritellään sopeutuvan dokumentin käsitettä ja dokumentin esittämistä siten, että sen automaattinen poiminta tietokannasta tai (TV-)lähetysvirrasta olisi mahdollista. Luonnollisen kielen keskeinen asema ja ongelmat dokumenttien löytämisessä tuodaan esiin. Samoin tarkastellaan lyhyesti agentteja dokumenttien suodattamisen välineinä ja hypermediaan perustuvia käyttöliittymiä niiden esittämisen välineinä. Lopuksi pohditaan tietokantajulkaisemisen teknisten muutosten vaikutuksia dokumenttien toimittamiseen ja kuluttamiseen sekä tiedotusvälineiden toimintaan.

Tietokantatoimittaminen ja -julkaiseminen

Parin viime vuosikymmenen nopean kehityksen tuloksena sanomalehdissä julkaistu ja sähköisissä viestimissä puhuttu teksti tuotetaan nykyisin valtaosaltaan elektronisessa muodossa. Elektronisia tekstejä arkistoidaan tekstitietokannoiksi, joiden sisältöä toimittajat voivat hyödyntää uusien uutisten kirjoittamisessa. Tänä päivänä elektronisoitumiskehitys on nopeaa myös kuva- ja ääniaineistojen osalta. Esim. sanomalehdissä luodaan elektronisia kuvatietokantoja, joihin talletetaan uutistoimistojen välittämät tai toimittajien ottamat digitoituidut valokuvat. Näin päästään tuottamaan ja taittamaan kaikki painettavien lehtien aineisto elektronisesti. Vastaavasti radiotoimittajien käyttämä ääniaineisto talletetaan digitaaliseen ja toimittajat voivat "taittaa" uutiset ja uutislähetykset elektronisella leikkauspöydällä esitysvalmiiksi. Samalla syntyy elektronisia äänidokumenttien tietokantoja. Kuitenkaan lopputuotteet, sanomalehdet tai radioiden uutislähetykset, eivät ole muodoltaan juuri muuttuneet saati elektronisoituneet.

Tietokannat ja tietoliikenneverkkojen kehitys kuitenkin tekevät elektronisen tiedonvälityksen mahdolliseksi. Tällöin lukija tilaisi itselleen elektronisen sanomalehden, jonka sisältö olisi sopeutettu paremmin lukijan henkilökohtaisiin tarpeisiin ja kiinnostuksen alueisiin kuin mikä on nykyisin tapana. Elektroninen sanomalehti voisi samalla tarjota tekstin ja elektronisten valokuvien lisäksi myös ääni- ja videoesityksiin perustuvia dokumentteja tai niiden osia. Sanomalehden taustalla oleva tietokanta tarjoaisi myös mahdollisuuden saada lisätietoa aihepiireistä, joiden käsittely elektronisessa lehdessä ei lukijalle riittäisi. Vastaavasti

myös esim. radion kuuntelija voisi tilata henkilökohtaisiin tarpeisiin sopeutetut uutiset omaan vastaanottimeensa.

Nämä visiot, tai pikemminkin lähitulevaisuuden mahdollisuudet, edellyttävät *tietokantajulkaisemista*, tiedonvälityksen dokumenttien julkaisemista tietokannan välityksellä. Julkaistavat dokumentit talletetaan tietokantaan, josta niitä voidaan vaihtelevin kriteerein poimia erilaisille lukijoille / katselijoille / kuuntelijoille eri jakelukanavissa (ml. perinteinen sanomalehti) esitettäväksi. Tietokantajulkaiseminen edellyttää *tietokantatoimittamista*, dokumenttien laatimista nimenomaan elektronista jakelua silmälläpitäen. Vain näin voidaan kuttakin toimitettua dokumenttia jakaa useammassa muodossa, moniin käyttötarkoituksiin ja -tilanteisiin. Tällöin dokumenttien sisältö ja rakenne pitää kuitenkin harkita uudelleen tällaista ympäristöä varten. Toimitustyön kannalta tietokanta on ensisijainen julkaisukanava, josta dokumentit leviävät muihin jakelukanaviin kuten valtakunnallisiin ja paikallisiin sanomalehtiin, paikallisradioihin, kaapelitelevisioon, internet-verkkoon jne., ja vastaanottiin kuten TV-vastaanotin, mikrotietokone, (korvalappu-) radio tai henkilökohtainen digitaalinen viestin (PDA, personal (tai portable) digital assistant).

Tällaista kehitystä edistävät teknologiset ja taloudelliset syyt, etenkin kustannus- ja julkaisuteknologian sekä tiedonvälitys- ja tietokoneteknologian integroituminen. Tämä kehitys vaikuttanee merkittävästi myös tiedonvälitysinstituutioiden organisaatioon ja toimintaan (Bair, 1993). Tässä artikkelissa esitellään tietokantajulkaisemiseen perustuvan joukkotiedotuksen teknologioita ja tarkastellaan niiden tarjomia mahdollisuuksia ja aiheuttamia ongelmia joukkotiedotuksessa.

Sopeutuvat dokumentit ja niiden löytäminen elektronisessa ympäristössä

Tiedonvälitys- ja tietokoneteknologian integroituminen johtaa elektroniseen ympäristöön, jossa käytetään tuhansia tietolähteitä (tietokantoja, TV- ja radiostudioita ja -asemia), satoja digitaalisia jakeluteitä (satoja digitaalisia TV- ja radiokanavia sekä tietokoneverkkoja) ja lukuisia erilaisia tiedon esitysvälineitä, kuten TV-ruudut, mikrotietokoneet, korvalappustereoradiot, datalazit, taskupuhelimet ja henkilökohtaiset digitaaliset viestimet. On periaatteessa järkevää pyrkiä siihen, että samaa dokumenttia (tai sen osia) voidaan jakaa mahdollisimman monin eri tavoin kuluttajille. Tähän liittyy sekä teknisiä että toimituksellisia ongelmia. Tarkastelen alla teknisiä ongelmia.

Tiedon esitysvälineet eroavat toisistaan niin median (video, audio, paperi) kuin kapasiteettinsa suhteen (esim. kuvaruudun tarkkuus ja koko, siirtonopeus, sivun koko ja painojälki). Tämän takia tietokantatoimittamiseen perustuvassa ympäristössä dokumentteja ei pidä jäsentää ja muotoilla vain paperille painamista varten vaan monipuolista elektronista käyttöä varten. Tällaisten dokumenttien tulee olla käyttöyhteyteen aktiivisesti sopeutuvia dokumentteja. MIT:n (Massachusetts Inst. of Technology) *News in the Future* -projektissa tutkitaan sopeutuvien dokumenttien tuottamista ja pyritään mm. seuraaviin dokumenttien ominaisuuksiin (Abrahamson & Bender, 1992):

- esityksen *skaalattavuus*: dokumentin esityksen tulee sopeutua annettuun esitystilaan, kuten kuvaruudulla olevan ikkunan koko, siirtokaistan leveyteen tiedonsiirrossa, tai laskentaresurssien saatavuuteen (esim. laitteiston kyky grafiikan esittämiseen);
- *esitystavan sopeutus käyttötilanteeseen*: esitysmateriaali, kuten video- tai audiomateriaali tulee

voida valita, samoin esitysnopeus ja vuorovaikutusmahdollisuudet eri käyttötilanteita (esim. olohuoneen TV-vastaanotin, salkkumikro puistossa, matkapuhelinyhteys juna-matkalla tai pyöräillessä) varten;

- sisällön sopeutus käyttötilanteeseen: sisällön tulee sopeutua painopisteiltään, näkökulmiltaan, kieleltään, jne. eri kohdeyleisöille (esim. ikäryhmät, kansallisuudet) ja eri tilanteisiin (esim. opetus- vs. viihdekäyttö).

Esityksen tulee sopeutua myös aikaan: esim. uutinen vanhenee uutislähetyksestä toiseen ja muuttaa samalla muotoaan. Myös tietokantaan talletetun uutisen tulisi vanheta siten, että koostettaessa uutiskokoelmaa ('lähetystä') automaattisesti tulisi vanhentuvien uutisten esityksen lyhetä. Käyttötilanteeseen sopeutumista on myös dokumentin tiivistysaste: 'sähköuutisissa' dokumentti esittäytyy lyhyempänä kuin 'päivän peilissä'.

Näin ollen tietokantajulkaiseminen edellyttää dokumenttien rakentamista siten, että niistä voidaan poimia eri esitysmuodoille sellaisenaan sopivat osat ('videokuva televisioon', 'videolähetysten ääniraita radioon') tai sopiviksi muunnettavissa olevat osat ('teksti syntesaattorin kautta puhelimeen'), eri käyttötilanteisiin sopivat osat ja erilaista tiivistämistä edustavat osat. Dokumenttien sopeuttaminen erilaisiin käyttötilanteisiin edellyttää mahdollisuutta dokumenttien osien valintaan ja uudelleenjärjestämiseen. Tätä helpottaa, jos kukin dokumentti kuuluu johonkin ennalta määriteltyyn dokumenttityyppiin, kuten pääkirjoitus, kuolinilmoitus, ulkomaan uutinen, valuuttakurssitaulukko tai auton myynti-ilmoitus sanomalehdessä tai mainospätkä, uutis-episodi tai TV-elokuva. Samaan tyyppiin kuuluvilla dokumenteilla on sama yleisrakenne ja useimmiten yhdenmukainen ulkoasu. Dokumenttien rakenteiden ja ulkoasujen kuvauksessa voidaan käyttää standardeja kuten ISO:n standardit SGML (Standard Generalized Markup Language) tai HyTime (esim. Heimburger, 1993).

Määritely rakenne tekee mahdolliseksi koostaa uusia dokumentteja olemassaolevista tai ainakin koota esikäsitelty aineisto sellaisia varten. Alla on tästä joitakin esimerkkejä:

- lehden sisällysluettelo ja kustantajan tuoteluettelo yms. hakemistot on helppo tuottaa automaattisesti rakenteisen dokumenttietokannan avulla
- katsaustyyppisten julkaisujen (kuten 'Bosnian sota sanoin ja kuvin', 'Sulo Aittoniemi ja somalit') perusmateriaali voidaan kerätä ja julkaisuprosessi (puoli)automatisoida rakenteisen dokumenttietokannan avulla
- jos dokumentti on kirjoitettu sanomalehdissä sovelletulla kärjellään seisovan pyramidin periaatteella, jolloin sitä voidaan lopusta melko vapaasti lyhentää, ja jos lyhennyskohdat on merkitty, voidaan tuottaa eri tasoisia uutistiivistelmiä lukijoiden/kuulijoiden mielenkiintoprofiilien mukaan.

Toinen tietokantajulkaiseminen aiheuttama ongelma koskee dokumenttien vastaanottoa. Elektronisessa ympäristössä tiedon käyttäjillä ei ole mahdollisuutta seurata koko tietovirtaa haluamansa tiedon saamiseksi. Heille syntyy tarve poimia automaattisesti talteen mahdollisesti kiinnostava tieto / dokumentti myöhemmin sopivana hetkenä tapahtuvaa käyttöä varten. Tämä edellyttää välineitä tiedon sisällön ja rakenteen hallintaan, jotta tieto voidaan poimia tietokannasta tai meneillään olevasta lähetysvirrasta.

Tekstiin perustuvasta tietovirrasta voidaan suodattaa kiinnostavia dokumentteja näiden sisältämien sanojen perusteella. Toistaiseksi kuvaan ja ääneen perustuvassa tietovirrassa tämä ei ole mahdollista. Kokeiluasteella on kuitenkin digitoitujen puhedokumenttien ääniperusteinen haku tietokannasta (Glavitsch & Schäuble, 1992). Kokeilussa oli mahdollis-

ta hakea radiossa esitetyt uutisdokumentteja ääneen tai tekstinä esitetyn kysymyksen avulla. Periaatteessa myös MIDI-koodattua musiikkia voitaisiin hakea siihen sisältyvien sävelkukujen nuottiesitysten avulla.

Toistaiseksi kuitenkin kuvaan ja ääneen perustuvaan tietovirtaan pitää liittää sisältöä kuvaava indeksointi, jotta vastaanotin voi aktivoitua ”nappaamaan” kiinnostavan episodin, esim. päivän uutis- ja ajankohtaislähetyksistä kaikki episodit, joissa poliitikot puhuvat seksipuhelimista. Tämä voidaan saavuttaa liittämällä myös kuvalähetykseen sisällönkoodaustieto, jonka perusteella vastaanotin voi poimia lähetyksen automaattisesti talteen. Tavallaan lähetykseen tarvitaan siis kolme ’raitaa’: kuvaraita, ääniraita ja koodiraita. Pentland & al. (1993) ovat kehittäneet videolähetyksen kuvakkeisiin perustuvan sisällönkuvaustyökulun, jossa kuva- ja ääniraidan ohella käytetään toistakymmentä koodiraitaa. Kukin raita on varattu jollekin sisällön piirteelle, kuten kameran liike ja kuvakulma, kuvaus aika, kuvauspaikka, kuvattujen henkilöiden käytöstapa (esim. istuminen / liike, katseen suunta). Kutakin sisältöpiirrettä kuvaava kuvake sijoitetaan raidalle aina siihen kohtaan, jossa vastaava sisältö esiintyy kuva- tai ääniraidalla. Vastaanotin voidaan nyt virittää poimimaan lähetyksestä vain ne jaksot, joihin liittyy tietty kuvake. Vaikka Pentlandin esittämät kuvakkeet koskevatkin varsin muodollisia piirteitä kuvavirrassa, ei mikään estä laajentamasta ideoita sisällöllisempään suuntaan. Monissa tapauksissa voidaan myös hyödyntää dokumentin elektronista tekstiä: ovathan esim. TV:n uutistenlukijan teksti tai ajankohtaisohjelman käsikirjoitus myös elektronisessa muodossa ja ne voidaan paikantaa kuvavirtaan.

Luonnollinen kieli hakuvälineenä

Tiedon hankinta elektronisten dokumenttietokantojen välityksellä voi tapahtua kolmella tavalla: (a) kuluttaja tilaa tietyn *tuotteen*, joka koostuu sovituin perustein valikoiduista ja muokatuista dokumenteista (vrt. sanomalehden tilaus), (b) kuluttaja hankkii itselleen *agentin* (’tekoälykäs’ tietokoneohjelma), joka suhteellisen itsenäisesti suodattaa käyttäjänsä suhteellisen pysyväismuotoisten tietotarpeiden kuvauksia vastaavia dokumentteja tietokannoista ja muista lähteistä, ja (c) kuluttaja tekee itse *kertaluontoisen haun* yhteen tai useampaan tietokantaan (esim. Ashford & Willet, 1988; Salton, 1989).

Sanomalehden tapaisia toimitettuja elektronisia tuotteita tarvitaan. Tietokannan välityksellä yksi ’lehtitalo’ voi kuitenkin tarjota suuremman valikoiman ja yksilöllisempiä sanomalehden elektronisia vastineita. Tällaisessa ’lehdessä’ tilaaja voisi valita useiden vakiotuotemallien välillä ja lisäksi virittää valitun mallin itselleen sopivaksi (esim. painottaa uutislajeja uutiskynnyksin ja uutisten esityspituuksin). Tuotteen tilaajan ei tarvitse ajatella elektronisten dokumenttien löytämiseen liittyviä ongelmia. Ne tulevat kuitenkin eteen muissa vaihtoehdoissa, joissa agentille joudutaan opettamaan kuvaus tiedontarpeista tai haku toteutetaan kokonaan itse. Haku elektronisessa dokumenttietokannassa voi perustua dokumentin tyyppiin (esim. kotimaan uutinen), rakenteeseen (esim. ingressi) ja sisältöön (esim. löytyy merkkipäivä ’aittoniem’ tai ’suti’) sekä ulkoasuun (esim. otsikon palstaleveys ja kirjaskoko).

Luonnollista kieltä oleva teksti on kuitenkin tiedonhaun keskeinen perusta. Merkittävä osa tiedosta on esitetty luonnollisella kielellä. Myös muun muotoista tietoa (kuten kuvia) haetaan tavallisesti siihen liitetyn tekstin (esim. kuvatekstin) avulla. Luonnollinen kieli tarjoaa parhaimman keinon dokumenttien sisältöön perustuvaan hakuun. Kielen monimuotoi-

suus aiheuttaa kuitenkin suuria ongelmia. Saman asian erilaisia ilmaisutapoja on runsaasti ja siksi lauseiden merkityksen automaattinen analyysi on hyvin vaikeaa. Esim. aihetta ”julki-sen talouden supistaminen” tarkasteltaessa supistaminen on voitu ilmaista mm. sanoilla supistaa, karsia, alasajo, hevoskuuri, kurittaa, keventää, ohentaa ja näiden kieliopillisilla johdannaisilla. Mutta myös monia muitakin asioita on ilmaistu samoilla sanoilla.

Luonnollisen kielen automaattisen käsittelyn vaiheita ovat (Carlson & Honkela, 1993; Karlsson, 1993):

- (i) Sananmuotojen morfologinen analyysi eli niiden palauttaminen perusmuotoon ja tietoihin taiputuspiirteistä.
- (ii) Morfologisten moniselitteisyyksien ratkonta eli disambigointi.
- (iii) Morfologisesti analysoitujen lauseiden syntaktinen analyysi eli jäsenitys, jossa määritellään kunkin sananmuodon syntaktinen tehtävä lauseessa.
- (iv) Lauseiden kirjaimellisen merkityksen tulkinta.
- (v) Lauseen kontekstimerkityksen tulkinta.

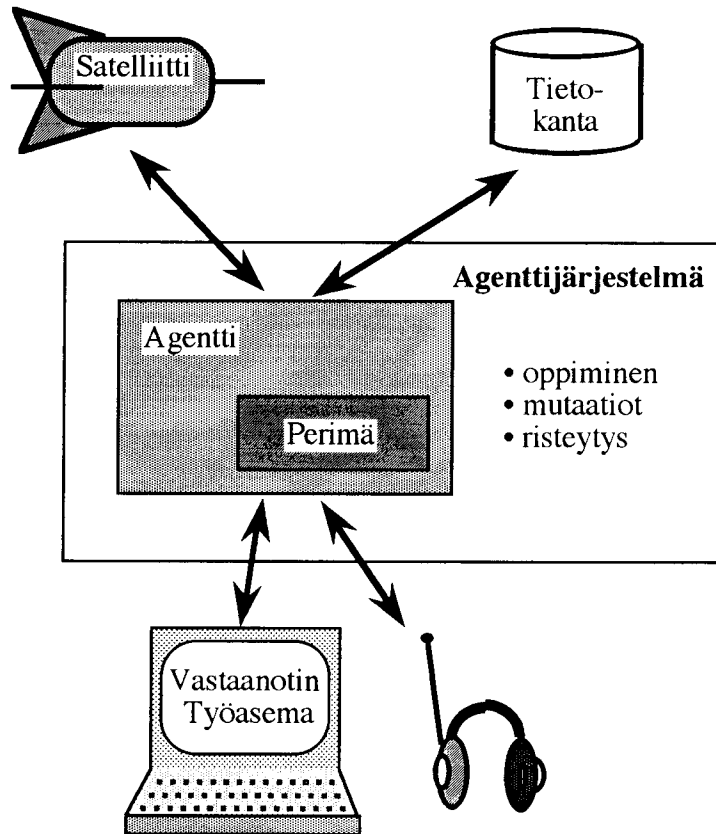
Luonnollisen kielen automaattisessa tulkinnessa on lukuisia ratkaisemattomia ongelmia varsinkin, jos tarkasteltavan kielen sanamäärää ei rajoiteta hyvin pieneksi (enintään satoja sanoja). Siksi lähivuosina tuskin voidaan ratkaista merkityksen tulkinnan ongelmaa rajoittamattomassa tekstissä automaattisin välinein. (Karlsson, 1993)

Tiedonhaussa tämä ei kuitenkaan ole edes välttämätöntä. Haun ongelmat eivät oleellisesti helpottuisi, vaikka minkä tahansa dokumenttien tekstin merkityksen automaattinen tulkinta onnistuisikin (vaiheet iv ja v). ”*Information goes beyond meaning*” — haun tarkoitus ei ole merkitysten haku vaan sellaisen informaation (käsiterakenteiden) haku, joka voi muokata vastaanottajan tietämystä (Ingwersen, 1992). Tällainen informaatio tulkitaan tekstistä eivätkä sen tulkin ehdot ole tekstissä itsessään vaan tulkitsijan omassa tilanteessa. On vaikeaa, ellei mahdotonta, kuvata ennen dokumenttien löytämistä, millaisia merkityksiä niiden teksteissä pitäisi lingvistisen tulkintaohjelman analysoimana olla, jotta ne olisivat lukijalleen hyödyllisiä.

Toisaalta jo pelkästään kolmen ensimmäisen käsittelyvaiheen menetelmiä tarvitaan tiedonhaussa, oikoluvussa, automaattisessa tavutuksessa, kielen kääntämisessä ja puheentunnistuksessa. Suomen kielen substantiivit voivat taipua noin 2000 eri muodossa (esim. ’töissänsäköhän’) ja verbit noin 12 000 muodossa. Morfologisten analyysiohjelmien avulla haku voidaan kuitenkin suorittaa sanojen perusmuodoilla (esim. ’työ’) ja näin sanojen taipumisen aiheuttamia hakuongelmia voidaan oleellisesti helpottaa (esim. Alkula & Honkela, 1992). Useiden mahdollisten ilmaisutapojen ongelmaa voidaan helpottaa sanastoilla, jotka kuvaavat ilmaisutapojen välisiä synonyymia- ja muita käsitteellisiä suhteita. Näin teksti voidaan usein löytää, vaikka sen hakija ja kirjoittaja käyttäisivät aivan eri ilmaisuja (Kristensen, 1993a-b).

Agentit — elektronisten dokumenttien suodatus

Agentti on dokumenttien suodatusväline (Kuva 1). Se on tietokoneohjelma, joka suhteellisen itsenäisesti seuraa määrättyä kohdetta (esim. TV-kanavaa tai tietokantaa), poimii siitä talteen annettuihin suodatuskuvauksiin täsmäivät dokumentit ja raportoi ne sovitulla tavalla käyttäjälle. Agentti sisältää perimän (genotyyppi, ohjeet) ja lisäksi ohjelmakoodin, joka



Kuva 1. Agentti dokumenttien suotimena

osaa käyttää tietolähteitä ja kanavia perimän määräämällä tavalla. Yksinkertaisen agentin perimä voi näyttää esim. seuraavalta (Seth & Maes, 1993):

```

uutisryhmä:   clari.sports.basketball
paikka:       boston, chicago, usa
lähde:        New York Times
avainsanat:   celtics, bulls, jordan, magic johnson
  
```

Tällainen agentti seuraisi New York Timesin uutisia clari.sports.basketball -uutisryhmästä tarkkailisi niistä avainsanoja celtics, bulls, jordan ja magic johnson. Näiden esiintyessä dokumentissa se poimittaisiin käyttäjälle esitettäväksi. Agentteja on kokeiltu informaation suodatuksen rohkaisevin tuloksin (Maes & Kozierok, 1993; Seth & Maes, 1993).

Agentin raportointiin liittyy oppimiskierros, jossa käyttäjä antaa palautetta löytyneiden dokumenttien hyödyllisyydestä. Agentti pyrkii täsmentämään suodatustehtävän kuvaustaan palautteen perusteella automaattisesti — lisää sellaisia dokumentteja, jotka arvioitiin hyödyllisiksi, ja vähemmän sellaisia, jotka arvioitiin hyödyttömissä. Agentin avainsanaraken-

netta voidaan muuttaa enemmän hyödyllisten dokumenttien kaltaiseksi ja vähemmän hyödyttömiä dokumenttien kaltaiseksi. Sopivia menetelmiä on kehitetty tiedonhaun tutkimuksessa (esim. Salton, 1989).

Käyttäjällä voi olla liuta agenteja eri tarkoituksiin, esim. eri aihepiirien tai eri tyyppisten tietolähteiden seuraamiseen. Oppimisen tuloksena eri agentit kehittyvät erilaisiksi ja myös niiden kokonaisyödyllisyys voi vaihdella. Samaa aihepiiriä seuraavien agenttien tulokset voidaan yhdistää ennen esittämistä käyttäjälle. Kun käyttäjä nyt antaa palautetta, voidaan suodatuskuvauksen täsmentämisen ohella agenteille antaa automaattisesti myös vahvuuspisteitä (kuten shakinpelaajille). Paljon hyödyllisiä dokumentteja löytävän agentin vahvuusluokka kasvaa ja niitä vähän löytävän agentin vahvuusluokka laskee. Vahva agentti saa raportoida enemmän löydöksiä käyttäjälle kuin heikko agentti. (Maes & Kozierok, 1993; Seth & Maes, 1993).

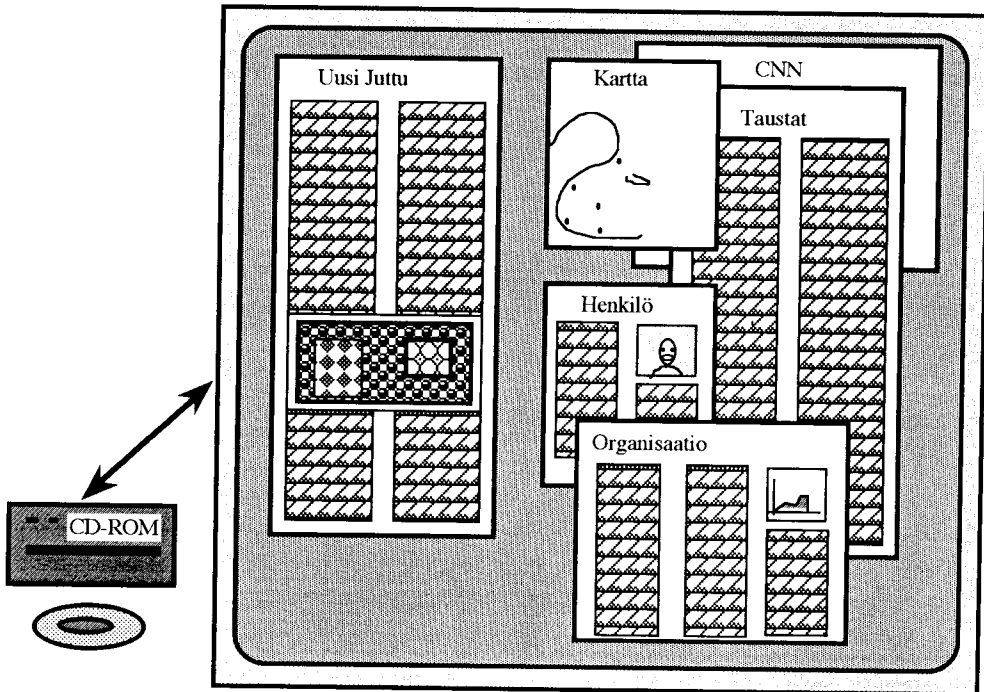
Agenttien oppimisen nopeuttamiseksi käytetään koneoppimisen menetelmiä. Ne perustuvat agenttisukupolviin ja agenttien muuntumisen biologiseen analogiaan. Vahvimmat agentit jäävät eloon ja heikoimmat karsiutuvat. Vahvimmat agentit saavat myös jälkeläisiä. Vain parhaimmiksi arvioidut perimän osat periytyvät jälkeläisille. Agenteille aiheutetaan mutaatioita (esim. vaihdetaan uutisryhmä toiseen, mutta ei liian kaukaiseen) ja niitä risteytetään (crossover) keskenään: kahden agentin jälkeläisten kesken voidaan esim. uutisryhmä ja muut perimän osat vaihtaa ristiin. (Maes & Kozierok, 1993; Seth & Maes, 1993).

Tämä teknologia ei ole kovin monimutkaista, vaikka saattaa tuntua kummalliselta biologisine vertauksineen. Avainsanojen ohella agenteja voi virittää tarkkailemaan esim. videodokumenttien koodiraitojen kuvakkeita (ks. yllä). Agenteja tarvitaan tiedon suodattamiseen informaatiotulvasta ja niitä päästään pian kokeilemaan Suomessakin. Varmaan joidenkin viestintäyritysten liikeidea tulee perustumaan ”hyvin koulutettujen agenttien” kauppaamiseen. Työtoverit ja ystävät varmaankin tulevat vaihtamaan hyviä agenteja keskenään.

Hypermediaan perustuvat käyttöliittymät

Hypermedian keskeinen idea on jäsentää informaatio solmuiksi ja niiden väliseksi yhteyksiksi (linkeiksi). Rakenteisissa hyperdokumenteissa voidaan määritellä solmutyyppejä niiden sisältämän informaation sisällön ja / tai muodon (teksti, ääni) perusteella. Lisäksi voidaan määritellä dokumenttityyppejä ja tyyppikohtaisesti se, millaisia solmutyyppejä dokumenttityyppiin liittyy (esim. Salminen & al., 1993). Perinteisesti hyperdokumentin kirjoittaja luo paitsi dokumentin solmut, myös niiden väliset linkit. Linkkien sanotaan edustavan ihmisten tapaa ajatella vapaasti assosioiden ja tekevän monet näkökulmat samaan dokumenttiin mahdollisiksi. Näihin piirteisiin liittyy julkisuudessakin paljon kritiikitöntä ylimyyntiä. Yksi ongelma-alue on linkkien muodostaminen. Dokumenttien käyttäjien vapaiden (mutta ehkä hyvinkin perusteltujen) assosiaatioiden arvaaminen ennalta on jokseenkin mahdotonta.

Frank Tompa (1992; Raymond & Tompa, 1988) on tutkinut suuren elektronisen sanakirjan ympäristössä automaattisen linkityksen menetelmiä. Linkitys perustuu sanakirjan tekstin rakenteeseen ja määriteltäisiin solmutyyppeihin. Rakenteen osia ovat esim. hakusana, sanaluokka, taivutus, etymologia, merkityskuvaukset, synonyymit, johdokset. Solmutyyppejä ovat esim. hakusanan yleisinformaatio (hakusana, sanaluokka, taivutus, merkitys-



Kuva 2. Kuvitteellinen uutisten hyperliittymä

vaihtoehdot), merkityskuvaukset, synonyymit ja viitesanat. Jokainen tietokannan sana on implisiittinen linkki useisiin mahdollisiin erityyppisiin solmuihin, joita voidaan esittää kuvaruudulla ao. tyypille tarkoitetuissa ikkunoissa.

Näitä ideoita voidaan soveltaa uutistietokantoihin. Kukin uutinen kuuluu johonkin uutislajiin (genre, esim. onnettomuus, sotatapahtuma, poliittisen henkilön toiminta). Kustakin uutislajista määritellään tavallisesti esiintyvät kategoriat sekä niiden esitystapa. Esimerkiksi onnettomuutta käsittelevä dokumentti kuvaa usein aikaa, paikkaa, onnettomuuden lajia, onnettomuuden syytä ja kulkua, aineellisia vahinkoja, inhimillisiä vahinkoja, ja pelastustoimenpiteitä. Poliittikan henkilön toimintaa käsittelevässä jutussa esiintyy ko. henkilö, aika, paikka, toiminnan laji, toiminnan kohteet (ihmisiä, organisaatioita, valtioita). Tällaisia jäsenyhsosia koskeva tutkimus on meneillään Tampereen yliopistossa.

Kuvan 2 esimerkissä on kuvitteellinen uutisten hyperliittymä. Sen esillä olevat solmutyypit perustuvat perussolmun (UutisJuttu) aktiiviseen dokumenttityyppiin (tässä poliittisen henkilön toiminta):

- UusiJuttu: agentin suodattama uutinen politiikan henkilön toiminnasta, joka esitetään

käyttäjän määräämässä laajuudessa; jutussa on kuvitteellisesti kaksi tekstipalstaa ja kuva;

- Kartta: näyttää jonkin uutisessa mainitun alueen kartan;
- Henkilö: esittää perushenkilötiedot uutisen kohteena olevasta politiikan henkilöstä tai muusta uutisessa mainitusta henkilöstä;
- Organisaatio: esittää organisaatiota koskevan informaation uutisessa mainitusta organisaatiosta (esim. henkilön taustaorganisaatio tai toiminnan kohde);
- Taustat: samaan uutisaiheeseen liittyvät vanhemmat uutiset tiivistettynä; voivat tulla joko uutispalvelusta tai olla käyttäjän tallettamia esim. viikon ajalta;
- CNN: uutiskanavan paikanpäältä otetuista videolähetyksistä talteen poimittu, jutun aihepiiriin liittyvä videomateriaali.

Muita mahdollisia solmutyyppejä voisi olla politiikan ja talouden sanakirja (koska jutuissa käytetään usein lukijalle outoja sanoja). Kun UusiJuttu -solmuun aktivoidaan seuraava juttu, jonka tyyppi on toinen, vaihtuvat muut solmutyypit automaattisesti jutun mukaisiksi. Kuvan CD-ROM on tietosanakirja, joka sisältää suhteellisen pysyvää informaatiota, kuten kartat, henkilö- ja organisaatitiedot. Ne voidaan jakaa kaikille tietyn uutispalvelun tilaajille taustamateriaaliksi, jolloin ko. informaatiota ei tarvitse jakaa jokaisen uutisen yhteydessä erikseen.

Linkitys solmujen välillä tapahtuu automaattisesti dokumentin käytön aikana. Jutun sanat toimivat linkkeinä taustajuttuihin: sanaa osoitettaessa Taustat-ikkuna esittää taustajuttuista ne katkelmat, jossa sana esiintyy. Aluksi Henkilö- ja Organisaatio-ikkunat voivat esittää jutun ensimmäisen henkilön/organisaation lisätietoja ja karttaikkuna näyttää yleiskartan juttuun liittyvästä paikasta (käyttäjä voi zoomata kohti kartan yksityiskohtia); Taustat-ikkuna näyttää viimeisimpien vanhojen juttujen tiivistelmät. Solmujen sisältö muuttuu jutun käsitteilyn mukana: juttua selattaessa tai sen osia aktivoitaessa, aktiivin alueen sisältö siirtää ikkunoita eteenpäin.

Sekä informaation haku että käyttö ovat vuorovaikuttaisia. Käyttäjä suuntaa hakuaan löytämänsä informaation perusteella. Käyttäjä käyttää dokumentteja vuorovaikuttaisesti hyppäämällä tarpeettomien osien yli ja syventää tarkastelua kiinnostavissa kohdissa. Työnjako 'kulutettavan' dokumentin luonnissa muuttuu siten, että informaation käyttäjä tulee aktiivisemmin osallistumaan haluamansa tuloksen rakentamiseen. Tavalliset kuluttajat tyytyvät varsin vähäiseen muokkaukseen kun taas ammattiyhteisöissä suodatus ja muokaus voi olla vaativaa.

Esimerkki oli kuvitteellinen ja perustui teksti- ja kuva/äänipohjaisen dokumentin käyttöön suureholla kuvaruudulla. Tällaisen käyttöliittymän toteuttaminen pienessä mittakaavassa ei edellytä periaatteellisten ongelmien ratkaisua. Suurempi kysymys onkin sen selvittäminen, millaiselta käyttöliittymän pitäisi näyttää ja mitä mahdollisuuksia tarjota, jotta se tyydyttäisi käyttäjiä. Tämä voidaan selvittää vain kokeilemalla prototyyppien avulla. Tällaisessa kokeilussa tarvitaan paitsi teknisten menetelmien asiantuntijoita ennenkaikkiaan toimittajia ja ajateltuja tulevia käyttäjiä.

Hyperdokumenttipiirteet tarjoava käyttöliittymä voidaan toteuttaa myös pelkästään äänidokumenttien ympäristössä. Chris Schmandt (1993a-b) MIT:n Media Labista on kehittänyt puhelimeen ja ääneen perustuvan uutisten ja muiden dokumenttien käyttöliittymän ja kokeillut sitä vaihtelevissa tilanteissa (puhelinpöydästä, matkapuhelimella autosta). Välitettä-

vät dokumentit voivat olla joko digitoitua puheetta tai tekstidokumentteja, jotka syntetisaattori esittää. Barry Arons (1991) on kokeillut pelkästään puheena välitettäviä hyperdokumentteja ja linkkien seuraamista puhutuin komennoin. Samoja ideoita on Tampereen yliopiston (Informaatiotutkimuksen laitos), Tampereen Teknillisen Korkeakoulun (Signaalinkäsittelyn laitos) ja Helsingin yliopiston (Yleisen kielitieteen laitos) ehdottamassa tutkimushankkeessa, jossa tutkittaisiin puhumalla (esim. matkapuhelimessa) esitetyn tiedonhakutehtävän suorittamista tekstitietokannassa (esim. sanomalehden uutisarkisto) ja vastauksen välittämistä puhuttuna kysyjälle. Tehtävä edellyttää puheentunnistusta, luonnollisen kielen käsitteitä, mahdollisten vastauksen osien suodatusta tietokannasta, vastauksen kokoonpanoa sekä puhesynteesiä.

Toimittaminen, tiedonvälitysinstituutiot ja julkisuus

Monet ovat esittäneet, että nykyiset tiedonvälitysinstituutiot tulevat elektronisen informaation lisääntymisen myötä tarpeettomiksi. Tiedon tarvitsijat voivat poimia haluamansa aineistot suoraan tarjolla olevista lähteistä eikä toimittajia tai uutistoimituksia tarvita tässä välikäsinä. Tämä näkemys perustuu kuitenkin ilmeiseen väärinkäsitykseen. Sanomalehdet tai TV:n uutistoimitukset eivät tule tarpeettomiksi, koska toimitukset ovat korvaamattomia tiedon suotimia ja organisoijia — ne tuottavat lisäarvoa, joka helpottaa elämää tai toimintaa tai antaa ainakin yhteisiä puheenaiheita. Mitä enemmän tietoa tuotetaan sitä enemmän tarvitaan sen organisointia ja suodatusta. Hyvin harvalla on käytännössä todella halua hankkia tietonsa suoraan tuhansista alkuperäislähteistä. (Bair, 1993).

Toimittamisen tärkeydestä ei kuitenkaan seuraa, että tiedonvälityksen nykyinen jako sähköisiin viestimiin ja sanomalehdistöön olisi pysyvä tai että nykyisten sanomalehtien asema ja toimintatapa olisivat pysyviä. Raja-aidat pikemminkin hämärtyvät. Tietokantajulkaisemisen riippumattomuus suurista painokoneista ja etäisyyksistä tekee mahdolliseksi purkaa esim. sanomalehtien koko organisaation: 'lehti' voi olla jatkuvasti päivittyvä tietokanta jossa-kin palvelinkoneessa. Jokainen 'lehteen' tuleva 'juttu' hankitaan tietoliikenneverkon välityksellä erikseen freelance-toimittajalta (vakituisia toimittajia ei ole), joka saa korvauksensa sen perusteella, kuinka moni kuluttaja vastaanottaa 'jutun' tai osan siitä jossakin muodossa. Toimittajan tekijänoikeus kenties hämärtyy tai katoaa, koska ei ole kokonaisuutta, jota se koskisi. Toimittaja tarvitsee uutta osaamista monikäyttöisten dokumenttien tuotannossa: aineiston keruussa elektronisista lähteistä ja monikäyttöisen sopeutuvan dokumentin laadinnassa.

Voidaan myös ajatella, että nykyisten tiedonvälitysinstituutioiden luoma julkisuus pirstoutuisi. Kansalaiset hankkisivat tietonsa niin monenlaisista eri lähteistä, ettei yhteistä julkisuutta syntyisi. Tämä on mahdollista, mutta vastavoimana toimii se, että julkisuutta tarvitaan. Se antaa kokemuksen yhteiskuntaan osallistumisesta. Kehityksen kulkua on vaikea ennustaa. Uskon kuitenkin, etteivät erilaiset teknologiat leviä vain sen takia, että ne tulevat tarjolle. Tullakseen laajaan käyttöön niiden tulee olla teknisesti vaivattomia, käytössä halpoja ja, mikä tärkeintä, kulttuurisesti yhteensopivia — antaa keino täyttää tarpeita, jotka ovat paljon monimuotoisempia kuin pelkästään sanoman saaminen vastaanottajan kuvaruudulle viehättävässä muodossa.

Lopuksi

Merkittävät teknologiset innovaatiot eivät ole välttämättömiä tässä kirjoituksessa kuvattujen tiedonvälityksen mahdollisuuksien toteuttamiseksi. Suurelta osalta mahdollisuuksien toteuttamisessa on kyse olemassa olevien teknologioiden (puhelin, televisio, tietokoneet) integroinnista ja soveltamisesta tietokantajulkaisemiseen. Tähän liittyy paitsi teknisiä, myös kaupallisia, ammatillisia, poliittisia ja kulttuurisia kysymyksiä:

- kaupallisia: kannattavien tuotteiden ideointi (millaista, kenelle, mihin käyttöön, millä hinnalla ?);
- ammatillisia: miten laaditaan tietokantadokumentteja, mikä on laatijan tekijänoikeus monimuotoiseen ja automaattisesti muokattuun tiedonjakeluun, millaiseksi kehittyä viestintäammattien sisältö;
- poliittisia: miten kehittyvät informaation saanti, osallistuminen, demokratia;
- kulttuurisia: teknologian ja kulttuuristen käytäntöjen supeutuneisuus puolin ja toisin (miten uudet välineet ja dokumenttimuodot sopeutuvat käyttöön ? miten uusia välineitä ja dokumenttimuotoja käytetään ? miten käyttö muuttaa muuta toimintaa ?), viihteen valta vs. henkilökohtainen osallistuminen, julkisuus.

Toivon, että käsillä olevan muutoksen teknologiapainotteinenkin kuvaaminen herättää toisenlaisen taustan omaavia tutkijoita ja ammattilaisia selvittämään kaupallisia, ammatillisia, poliittisia ja kulttuurisia kysymyksiä. Teknologioiden integroituminen edellyttää myös tutkimuksen integroitumista akselilla informaatiotutkimus — tiedotusoppi — tietojenkäsittelyoppi.

Tiedonvälityksen teknologia on muuttunut vuosikymmenten mittaan voimakkaasti. Muutos varmasti jatkuu edelleen ja tarjoaa mahdollisuuksia rahan, työpaikkojen, vallan ja kulttuurihegemonian uusjakoon rahvaan, tiedotuksen ammattilaisten ja ammattijärjestöjen, alan yritysten ja valtioiden kesken. Tietämisen tason tai sivistyksen kanssa näillä muutoksilla ei liene paljon tekemistä. Kauppa kuitenkin käy, varsinkin viihteen massamarkkinoilla. Pään voi varmasti panna pensaaseen ja sanoa ”Ei ei, tällaista emme tarvitse / halua / salli”. Väki kuitenkin äänestää kaukosäätimellään tai PDA:llaan.

KIRJALLISUUS

- Abrahamson, N. & Bender, W. (1992) Context-Sensitive Multimedia. In: Proc. of the SPIE, vol. 1785. (Reprint, MIT/Media Laboratory, 1993).
- Alkula, R. & Honkela, T. (1992) Tekstin tallennus- ja hakumenetelmien kehittäminen suomen kielen tulkinohjelmien avulla. Espoo: Valtion teknillinen tutkimuskeskus, Julkaisuja 765.
- Arons, B. (1991) HyperSpeech: Navigating in Speech-Only Hypermedia. In Proc. 1991 Hypertext Conference. New York, NY, ACM.
- Ashford, J.H. & Willett, P. (1988) Text Retrieval and Document Databases. Chartwell-Bratt, Lund.
- Bair, P. (1993) New Media and Newspaper Publishing. In: Proc. EP International '93, IEPRC XIIIth Annual Conference, Oslo, June 1993.
- Carlson, Lauri & Honkela, Timo (1993) Luonnollisen kielen käsittely. Teoksessa: Hyvönen, E. & Karanta, I. & Syrjänen, M., Tekoälyn ensyklopedia. Hki: Gaudeamus, ss. 47-52.
- Glavitsch, U. & Schäuble, P. (1992) A System for retrieving speech documents. Belkin, N. & Ingwersen, P. & Mark Pejtersen, A.-L. (eds.) Proc. of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Special Issue of SIGIR Forum, June 21-24, 1992, pp. 168-176.
- Heimbürger, Anneli (1993) Teknisen dokumentaation kehittäminen hypertext- ja CD-ROM -tekniikan

- avulla. Espoo: Valtion teknillinen tutkimuskeskus, VTT Tiedotteita 1489. 104 s.
- Ingwersen, P. (1992) Information retrieval interaction. London: Taylor Graham.
- Karlsson, Fred (1993) Kielitiede. Teoksessa: Hyvönen, E. & Karanta, I. & Syrjänen, M., Tekoälyn ensyklopedia. Hki: Gaudeamus, ss. 47-52.
- Kristensen, J. (1993a) Hakutesauruksen vaikutus sanahakuihin. Kirjastotiede ja informatiikka 12(3):95-104.
- Kristensen, J. (1993b) Expanding End-User's Query Statements for Free Text Searching with a Search-Aid Thesaurus. Information Processing & Management 29(6):733-745.
- Maes, P. & Kozierok, R. (1993) Learning Interface Agents. To appear in: Proceedings of the American Association for Artificial Intelligence 1993 Conference. (Preprint, MIT/Media Laboratory, 1993).
- Pentland, A. & al. (1993) The BT/MIT Project on Advanced Image Tools for Telecommunication: An overview. MIT Media Laboratory Perceptual Computing Group Technical Report #212
- Raymond, D.R. & Tompa, F. Wm. (1988) Hypertext and the Oxford English Dictionary. Communications of the ACM 31(7):871-878.
- Salminen, A. & Tague-Sutcliffe, J. & McClellan, C. (1993) From Text to Hypertext by Indexing. ACM Transactions on Information Systems, 1994 (to appear).
- Salton, G. (1989) Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley, Reading, MA,.
- Schmandt, C. (1993a) From Desktop Audio to Mobile Access: Opportunities for Voice in Computing. In: Hertson, H.R. & Hix, D. (eds.) Advances in Human-Computer Interaction, Vol 4. Noewood, NJ: Ablex, pp. 251-283.
- Schmandt, C. (1993b) Phoneshell: the Telephone as Computer Terminal. Proc. ACM Multimedia Conference, Anaheim, CA, August 1993.
- Seth, B. & Maes, P. (1993) Evolving Agents for Personalized Information Filtering. Proc. of the 9th Conference on Artificial Intelligence for Applications. (Preprint, MIT/Media Laboratory).
- Tompa, F. (1992) The New OED Project. Esitelmä tietojenkäsittelytieteen kesäkoulussa, Selänpohjassa, 14.- 17.6.1992.