

## **Idean kontekstia kuvaava metatieto**

Olli Niinivaara

Helsinki 2.10.2002

Pro gradu –tutkielma

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

Tiedekunta/Osasto <b>Matemaattis-luonnontieteellinen</b>		Laitos – Institution <b>Tietojenkäsittelytieteen laitos</b>	
Tekijä – Författare <b>Olli Niinivaara</b>			
Työn nimi – Arbetets titel <b>Idean kontekstia kuvaava metatieto</b>			
Oppiaine – Läroämne <b>Tietojenkäsittelytiede</b>			
Työn laji – Arbetets art <b>Pro gradu –tutkielma</b>		Aika – Datum <b>2.10.2002</b>	Sivumäärä – Sidoantal <b>82</b>
Tiivistelmä – Referat  <p>Tutkielmassa esitellään idean kontekstin kuvaaminen keinona tehostaa ideoiden välittymistä. Kontekstitieto kuvataan dokumentteihin liittyvänä metatietona, jota hallitaan dokumenteista riippumattomissa metatietokannoissa. Päämääränä pidetään sellaista idean kontekstin kuvausta, joka on riittävän ilmaisuvoimainen, mutta jonka luominen ei aseta järjestelmän käyttäjille ylivoimaista työtaakkaa.</p> <p>Tiedon välittyminen nähdään prosessina, johon perustuen idean konteksti jaetaan tuotokontekstiin, julkaisukontekstiin ja käyttökontekstiin. Tähän jakoon perustuen käsitellään metatiedon muodostaminen ja sisältö yksityiskohtaisesti yksittäisten metatietotietueen attribuuttien tasolla.</p> <p>Kontekstitiedon käyttökohteista tarkastellaan kontekstin visualisointia informaation visualisoinnin tekniikoihin perustuen, idean arvon mittaamista bibliometrisiä menetelmiä kehittämällä ja automaattista ideoiden valintaa tiedon suodatuksen menetelmien ja digitaalisten assistenttien avulla.</p> <p>CR-luokitus:</p> <p>H.1.0 [Models and principles]: General  H.1.0 [Models and principles]: Systems and Information Theory - <i>Value of information</i>  H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval  H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia - <i>Theory</i>  I.7.4 [Document and Text Processing]: Electronic Publishing</p>			
Avainsanat – Nyckelord <b>konteksti, idea, dokumentti, hypermedia, metatieto, tiedon välitys, Internet</b>			
Säilytyspaikka – Förvaringställe <b>Tietojenkäsittelytieteen laitoksen kirjasto, sarjanumero C-2002-59</b>			
Muita tietoja – Övriga uppgifter			

## Sisältö

1 Johdanto.....	1
2 Taustaa.....	5
2.1 Metatieto.....	5
2.2 Hypermedia.....	8
2.3 Lähdeviittausten hyödyntäminen.....	10
3 Kontekstin malli.....	13
3.1 Kontekstin käsite.....	13
3.2 Ideoiden välittymisprosessi.....	15
3.3 Avoin arkkitehtuuri.....	21
4 Kontekstin kuvaaminen.....	27
4.1 Idean kuvaaminen.....	27
4.2 Ideoiden väliset suhteet.....	31
4.3 Julkaisukonteksti.....	38
4.4 Metatietokuvauksen sisältö.....	44
5 Kontekstin käyttö.....	49
5.1 Tiedon tarvitsijan näkökulma.....	49
5.2 Kontekstin visualisointi.....	51
5.3 Idean arvon mittaaminen.....	55
5.4 Automaattinen tiedon valinta.....	60
6 Yhteenveto.....	66
Lähteet.....	70

## 1 Johdanto

Tiedon välittymistä helpottavia keksintöjä voidaan pitää koko ihmiskunnan mittakaavassakin hyvin merkityksellisinä. Esimerkkeinä mainittakoon kieli, kirjoitustaito, kirjapainotaito, kirjastot, puhelin ja Internet. Näihin päiviin asti tiedon välityksen tekniset ongelmat, kuten tiedon eri esitysmuotojen tallentamisen vaikeus, esitysten kopioinnin kalleus ja tiedonsiirron hitaus ovat toimineet tiedon välittymisen rajoittajina. Näiden esteiden poistuessa on rinnalle noussut uusi, saatavilla olevan tiedon määrästä johtuva ongelma. Tätä ongelmaa voisi kutsua tiedon valinnan ongelmaksi. Kun saatavilla olevan tiedon määrä kasvaa, tietyssä tilanteessa tarpeellisen tiedon määrä suhteessa muuhun saatavilla olevaan tietoon pienenee. Ongelma näkyy ideoiden tuottajien osalta siten, että arvokkaatkin ideat ovat vaarassa hukkua muuhun tietomassaan. Tiedon tarvitsijoiden kannalta ongelma näkyy siten, että arvokkaiden tietojen erottamiseksi tiedon valintaan joudutaan käyttämään yhä enemmän ja enemmän resursseja.

Tiedon valinnan ongelmaa voidaan ratkoa monilla eri tavoilla. Saatavilla olevan tiedon määrää voidaan vähentää esimerkiksi käyttämällä tehottomampia tiedonvälitysjärjestelmiä tai rajoittamalla tiedon julkaisemista tai saatavuutta. Tiedon tarvetta voidaan vähentää keskittymällä suppeampien tiedon alueiden hallintaan. Tiedon valintaa voidaan helpottaa siirtämällä tiedon valintatyötä useille tahoille, kuten erillisille tiedon välittäjille. Tässä tutkielmassa keskitytään yhden ratkaisuvaihtoehdon esittelyyn. Tämä on tiedon valinnan tukeminen automaattisten, tietojenkäsittelyllisten apuvälineiden avulla.

Tutkielmassa keskitytään ideoiden valinnan helpottamiseen. Ideat ovat tietoa, joka selittää jotain yleistä asiantilaa. Selitykset ovat väitteiden, eivät varmojen totuuksien asemassa – siksi ideat kaipaavat yleensä perusteluja ollakseen käyttökelpoisia. Ideoiden tarkoituksena on toimia ajattelun apuvälineinä. Ideoita tarvitaan silloin, kun on ymmärrettävä tai hahmotettava jokin asia tai ratkaistava jokin ongelma. Sekä ideoiden luominen että niiden käyttö vaativat ajatustyötä, kuten luovuutta, tulkintaa ja ymmärtämistä. Esimerkkejä tiedosta, jotka tässä luetaan kuuluvaksi ideoihin, ovat

teoriat, mallit, keksinnöt, analyysit ja raportit. Ideat ovat sitä arvokkaampia, mitä enemmän ne helpottavat asioiden ymmärtämistä, mitä tehokkaampia ongelmien ratkaisemisen apuvälineitä ne ovat ja mitä enemmän niistä on hyötyä uusien hyödyllisten ideoiden tuottamisessa. Ideoiden valinta on tehokasta silloin, kun tiedon tarvitsija pystyy mahdollisimman vähillä resursseilla poimimaan oman tiedon tarpeensa kannalta arvokkaimmat ja vain arvokkaimmat ideat kaikkien periaatteessa saatavilla olevien ideoiden joukosta.

Yksi paljon käytetty menetelmä tiedon löytämisen ja valinnan helpottamiseksi on kuvata tiedon sisältö tiiviissä muodossa esimerkiksi otsikon, asiasanojen, luokituksen tai tiivistelmän avulla. Menetelmä on tehokas, mutta ideoita tarvittaessa usein riittämätön. Idean arvo riippuu huomattavassa määrin useista sisällön ulkopuolisista seikoista kuten siitä, mihin tietoihin idea perustuu, miten ymmärrettävästi idea on esitetty, mistä näkökulmasta idea on tulkittava tai millaiseksi idean hyödyllisyys tai merkitys on käytännössä osoittautunut. Vaikka idean sisältö ei muutu, idean arvo voi hyvinkin muuttua sitä mukaa, kun ideaa ja idean käsittelemää aihetta koskeva tieto muuttuu. Tutkielmassa tarkastellaankin mahdollisuuksia mallintaa, kuvata, hallita ja hyödyntää tällaista idean arvon kannalta merkittävää tietoa, jota ei voi päätellä pelkästään idean tietosisällön perusteella. Tarkoituksena ei ole korvata sisällön kuvaukseen perustuvia menetelmiä, vaan kehittää niiden rinnalle ratkaisuja, jotka oleellisesti lisäävät tiedon tarvitsijan mahdollisuuksia arvioida ja valita tietoa.

Perusongelma on idean ympäristöä kuvaavan aineiston määrittäminen. Tätä aineistoa kutsutaan jatkossa idean kontekstia kuvaavaksi tiedoksi tai lyhyesti vain kontekstitiedoksi. Koska idean ympäristöön kuuluu käytännössä suunnattoman paljon tietoa, koko kontekstia ei voida kuvata, vaan konteksti on rajattava määrittelemällä se. Määritelmän tulee olla kaikille ideoille yhteinen, jotta kuvaukset ja siten ideat ovat keskenään vertailukelpoisia. Kuvauksen tulee olla riittävän yksinkertainen ja helppokäyttöinen, jotta sitä osataan hyödyntää. Toisaalta sen tulee olla riittävän ilmaisuvoimainen, jotta siitä olisi hyötyä. Kuvauksen merkityksen tulee olla määritelty niin yksityiskohtaisesti, että sitä voidaan käyttää automaattisten tiedon päättely- ja johtamismenetelmien syöteaineistona.

Kontekstitiedon hyödyllisyyttä tarkastellaan ideoiden tuottajien ja ideoiden tarvitsijoiden näkökulmista, ei esimerkiksi tiedon välittäjien, kuten kustantajien tai kirjastojen näkökulmista. Tuottajien kohdalla oletetaan, että ideoiden julkaisemista järjestelmässä motivoi se, että arvokkaat ideat voidaan erottaa arvottomista. Toisin sanoen, vaikka tiedon tuottajan välillisenä motivaationa saattaisivatkin olla esimerkiksi taloudelliset intressit, urakehitys tai 'brändin' luonti, niin välittömäksi motivaatioksi oletetaan se, että arvokkaita ideoita kannattaa julkaista järjestelmässä, jossa ne saavat arvoisensa huomion. Tiedon tarvitsijoiden näkökulmasta taas jonkinlaisena malliesimerkkinä voidaan pitää henkilöä, joka tarvitsee työkseen ideoita, työn laatu riippuu valittavien ideoiden laadusta ja tiedon tarve on laaja-alainen. Tällaisia tiedon tarvitsijoita löytyy tiedemaailmasta, politiikan saralta, yritysten johdosta, suunnittelu- ja tuotekehitystehtävistä jne. Laaja-alainen tiedon tarve voidaan määrittellä sellaiseksi, jossa ideoita tarvitaan useista tiedon lähteistä yhteisörajat ja tiedonalat ylittäen, jolloin tiedon valinta perinteisempiin menetelmiin tukeutuen on usein vaikeaa.

Tärkein kontekstitiedon määrittelemistä ohjaava periaate on se, että määritelmän on tarjottava keinot tietokoneiden ja tietoverkon tietojenkäsittelyllisen potentiaalin tehokkaaseen hyödyntämiseen. Koska Internet tiedonvälityksen teknisenä perusinfrastruktuurina on avoin, eivätkä tietokoneet vaadi tekemästään työstä taloudellisia tai muitakaan etuja, voi myös niihin perustuva järjestelmä olla avoin. Kun idean kontekstia kuvaava tieto esitetään digitaalisessa muodossa, voi kuvaus olla dynaaminen eli sitä voidaan tarpeen mukaan muuttaa tai kasvattaa. Tietokoneiden ja kehittyneiden käyttöliittymien avulla ideoita voidaan tehokkaasti vertailla, järjestää ja luokitella tiedon tarvitsija- ja tiedon tarvekohtaisesti. Kun kontekstitiedon tietosisältö on määritelty riittävän tarkasti, voidaan siitä muodostaa automaattisesti päätelmiä ja johtaa idean valinnan kannalta arvokasta tietoa ideasta.

Luvussa 2 esitellään kontekstitiedon hallinnan kannalta keskeiset menetelmät, jotka ovat metatieto kontekstin tallennuskeinona, hypermedia suhteiden ja assosiaatioiden hallintamenetelmänä sekä lähdeviittaukset ideoiden välisiä suhteita kuvaavana syöteaineistona. Luvussa 3 määritellään idean konteksti esittelemällä idean tuotto-, julkaisu- ja käyttökontekstit, kuvataan yleisellä tasolla ideoiden välittymisen prosessia sekä kiinnitetään konteksti osaksi tiedonvälitysjärjestelmää määrittelemällä avoin

arkkitehtuuri, jossa kontekstietoa hallitaan dokumenteista riippumattomissa metatietotietueissa. Luvussa 4 esitellään metatietomalli, joka soveltuu ideoiden kontekstia kuvaavan tiedon hallintaan. Metatietotietueiden sisältö määritellään yksittäisten attribuuttien tasolla. Luvussa 5 käsitellään kontekstiedon käyttöä ideoiden ja muiden resurssien etsimisen ja arvioinnin tukena. Käyttökohteista tarkastellaan kontekstin visualisointia informaation visualisoinnin tekniikoihin perustuen, idean arvon mittaamista bibliometrisiä menetelmiä kehittämällä ja automaattista ideoiden valintaa tiedon suodatuksen menetelmien ja digitaalisten assistenttien avulla. Luku 6 on yhteenveto.

## 2 Taustaa

### 2.1 Metatieto

Tiedonvälitysjärjestelmissä käsiteltävä tieto koostuu tietoalkioista eli resursseista. Resurssit ovat yksiköitä, jotka järjestelmässä voidaan tunnistaa omiksi kokonaisuuksikseen. Järjestelmässä käytetystä tiedon välityksen mallista riippuen resurssit voivat olla esimerkiksi sivuja, sivustoja, kuvia, viestejä, dokumentteja, tietovarastoja, palvelimia tai kanavia. Metatieto (metadata) on automaattisesti käsiteltävissä olevaa tietoa, jonka tarkoituksena on kuvata resursseja [DeH98, Ber97, Dil01]. Metatieto lisää resurssista saatavilla olevan tiedon määrää. Itse resurssit eivät ole metatietoa, eikä resursseista automaattisesti johdettua tietoakaan ole tapana pitää metatietona.

Jotta metatietoa voitaisiin käsitellä automaattisesti, metatiedon esitystavan (syntaksin) ja merkityksen (semantiikan) tulee olla määriteltyjä. Esitystavan määritelmää voidaan kutsua metatietoformaatiksi ja tietosisällön määritelmää metatietomalliksi. Metatietoformaatti on tekninen määritelmä, joka määrittelee metatiedon esitystavan konkreettisenä järjestelmässä hallittavana datana. Metatietomalli puolestaan on abstraktimpi malli, joka määrittelee metatiedon merkityksen. Mallia voitaisiin kutsua myös ontologiaksi [Obi01], sikäli kun se kiinnittää ne peruskäsitteet, joihin tiedonvälitysjärjestelmän toiminta perustuu. Tarkastellaan seuraavaksi tarkemmin yhtä olemassaolevaa metatietomallia.

Dublin Core (DC) on kansainvälisen Dublin Core Metadata Initiative (DCMI) –ryhmän kehittämä metatietomalli [Hil01, DeW02]. DC:n tarkoituksena on toimia yleiskäyttöisenä metatietostandardina, joka määrittelee muita kehitettyjä ja kehitettäviä metatietostandardeja yhdistävän perussemantiikan. DC:ssä käsiteltäviä resursseja kutsutaan yksinkertaisesti vain resursseiksi. Resurssin sisältöä tai merkitystä ei ole tarkemmin määritelty tai rajattu. Resurssia voidaan kuvata viidentoista eri attribuutin avulla, joita DC –mallissa kutsutaan elementeiksi. Nämä 15 elementtiä ovat seuraavat [Dub99, Hel00]:



- 1. Nimeke:** Tekijän tai julkaisijan antama resurssin nimi.
- 2. Tekijä:** Henkilö tai yhteisö, joka on ensisijaisesti vastuussa resurssin intellektuaalisesta sisällöstä.
- 3. Aihe:** Resurssin sisällön aihealueen kuvaus, yleensä asiasanoilla tai fraaseilla esitettynä.
- 4. Kuvaus:** Resurssin sisällön vapaamuotoinen kuvaus. (Kuvaus voi sisältää esimerkiksi tekstidokumentin tiivistelmän tai sisällysluettelon.)
- 5. Julkaisija:** Organisaatio, henkilö tai palvelu, joka on julkaissut tai asettanut resurssin käytettäväksi.
- 6. Muu tekijä:** Henkilö, organisaatio tai palvelu, joka Tekijä-kentässä mainittujen tekijöiden lisäksi on osallistunut merkittävästi resurssin luomiseen, mutta jonka panos on ollut toissijainen verrattuna ns. päävastuullisen tekijän osuuteen. (Muu tekijä voi olla esimerkiksi toimittaja, kuvittaja, kääntäjä, sisällöntuotantopalvelu.)
- 7. Päivämäärä:** Päivämäärä, joka liittyy resurssin johonkin elinkaaren vaiheeseen. (Yleensä Päivämäärä ilmaisee tallenteen julkistamisajan, luontiajan, voimassaoloajan tai ajanjakson, jolloin resurssi on käytettävissä.)
- 8. Laji:** Resurssin kirjallisuuslaji, luonne tai genre. (Resurssin lajina voi olla esimerkiksi kotisivu, romaani, runo, tekninen raportti, essee tai sanakirja.)
- 9. Formaatti:** Resurssin fyysinen tai digitaalinen ilmiasu.
- 10. Identifikaatiotunnus:** Tunnus, joka identifioi resurssin yksiselitteisesti.
- 11. Lähde:** Tunnus, joka identifioi yksiselitteisesti resurssin, johon kuvailtava resurssi perustuu kokonaan tai osittain.
- 12. Kieli:** Resurssin kieli.
- 13. Suhde:** Viittaus muuhun resurssiin, joka liittyy kuvailtavaan resurssiin.
- 14. Kate:** Resurssin sisällön alueellinen tai ajallinen kate.
- 15. Tekijänoikeudet:** Resurssiin liittyvät oikeudet.

Minkään elementin käyttö ei ole pakollista. Jokaiseen elementtiin voi liittyä useampia kuin yksi arvo. Arvoihin voi liittyä niiden merkityksen tarkenteita, jotka voivat olla DCMI:n suosituksia [Dub00] tai muita DC:n semantiikan tarkennuksia.

DC ei metatietomallina kiinnitä sitä, millä metatietoformaateilla elementit ja arvot tulee esittää. Esityksen perusrakenne on kuitenkin seuraava. Yhtä resurssia vastaa yksi

metatietokuvaus. Metatietokuvaus esitetään metatietotietueena, joka koostuu 'elementti = arvo' –muotoisten sijoituslauseiden luettelosta. Arvojen tulee olla merkkijonoja tai merkkijonoluetteloita. Käytettävästä metatietoformaattista riippuu, miten arvot tulee koodata merkkijonoiksi. DC:n esitystapa on määritelty useille metatietoformaateille. DC:n metatietomalli voidaan myös esittää yleisemmän Resource Description Framework (RDF) –mallin avulla, joten kaikki sille kehitetyt ja kehitettävät metatietoformaatit soveltuvat myös DC:n esittämiseen [Mil98, KoS02, BrG02]. Tässä tutkielmassa esitettävä metatietomalli kehitetään siten, että se ei aseta metatietoformaatile suurempia vaatimuksia kuin DC. Näin menetellen kaikki DC:lle ja siten myös RDF:lle soveltuvat metatietoformaatit ovat käyttökelpoisia. Jatkossa keskitytäänkin metatietomallin tarkasteluun ja metatietoformaatteihin liittyvät kysymykset ohitetaan.

Dublin Core on kehitetty kaikentyyppisiä resursseja koskevan metatiedon malliksi. Yksi sen suunnittelun keskeisistä lähtökohdista on ollut säilyttää yhteensopivuus olemassaolevien metatietostandardien, erityisesti kirjastojen kortistoissaan käyttämien standardien kanssa. Tässä tutkielmassa puolestaan kehitetään metatietomallia, joka soveltuu ideoiden välityksen, erityisesti ideoiden valinnan, tehostamiseen. Tarkoituksena ei ole säilyttää yhteensopivuutta olemassaoleviin malleihin, vaan kehittää malli, jonka avulla tietokoneiden ja tietoverkkojen tietojenkäsittelyllisiä ominaisuuksia voidaan mahdollisimman tehokkaasti hyödyntää. DC:hen ja yleisesti perinteisempiin tiedon kuvaustapoihin verrattuna keskeisiä ominaisuuksia ovat tällöin erilaisten resurssityyppien käyttö, resurssien välisten suhteiden ja assosiaatioiden kuvaaminen ja näiden kuvausten monipuolinen ja tehokas hyödyntäminen sekä dynaamisesti muuttuva resurssien kuvaus.

## 2.2 Hypermedia

Hypermedialla (hypertekstillä) tarkoitetaan tekniikoita, joiden avulla resurssien välisiä suhteita voidaan kuvata ja esittää [Con87]. Yksittäistä resurssien välillä vallitsevaa suhdetta kutsutaan linkiksi. Hypermediajärjestelmän tarkoitus riippuu siitä, millaisia suhteita järjestelmässä kuvataan.

Ensimmäisenä hypermedian kuvauksena pidetään yleisesti Bushin artikkelia vuodelta 1945 [Bus45]. Bush toteaa, että yhä kasvavan ja monimutkaistuvan tietomäärän hallitsemisessa on ongelmana tiedon valinnan vaikeus. Ihminen hahmottaa tietoa assosiaatioiden avulla, joten luonteva tapa helpottaa tiedon valintaa olisi eksplisiittisesti kuvata tietojen välillä vallitsevia suhteita. Bushin näkökulmasta hypermedian tarkoitus on kuvata tietojen välisiä suhteita siten, että tiettyyn aiheeseen liittyvät tiedot ovat helposti löydettävissä assosiaatiolinkkien muodostamia polkuja seuraamalla.

Toisen näkökulman hypermedian käytölle esittelee Engelbart raportissaan vuodelta 1962 [Eng62]. Raportissa hahmotellaan tapoja tehostaa ihmisten ongelmanratkaisukykyä erityisesti tietokoneiden avulla. Engelbart näkee assosiaatioiden kuvaamisen tehokkaana tapana esittää ja jäsenellä monimutkaiseen ongelmaan liittyviä ja ongelman ratkaisuprosessin aikana syntyviä käsitteitä ja ajatuksia. Tästä näkökulmasta hypermediatekniikoiden tarkoituksena on helpottaa tiedon ymmärtämistä tietojen välisiä pragmaattisia, tiedon käyttöä ja tiedon merkittävyyttä kuvaavia suhteita kuvaamalla. Tämän tyyppisiä linkkejä kutsutaan usein argumenttilinkeiksi tai retorisisiksi linkeiksi.

Kolmas yleisesti tunnettu hypermedian käyttökohde on sen käyttö tiedon esitysmuotona. Tämän näkökulma hypermediaan on erityisesti Nelsonin kehittämä [Nel99]. Kun esityksiä voidaan linkkien avulla kytkeä toisiinsa, voidaan uusia esityksiä luoda koostamalla niitä muista esityksistä. Tästä näkökulmasta hypermedian tarkoituksena on helpottaa tiedon esittämistä ja tarjota vaihtoehto perinteiselle lineaariselle tekstille.

Ideoiden kontekstin kuvaamisen kannalta esitetyistä näkökulmista ovat relevantteja Bushin näkökulma, eli ideoiden etsimisen ja valinnan helpottaminen ideoiden välisiä suhteita kuvaamalla, sekä etenkin Engelbartin näkökulma, eli ideoiden arvioinnin ja ymmärtämisen helpottaminen ideoiden välisiä pragmaattisia yhteyksiä kuvaamalla. Sen sijaan hypermediaa ideoiden esittämissyntaksina ei jatkossa käsitellä. Ideat voivat olla esitettyjä hypermediamuotoisina, mutta niiden ei tarvitse olla. Kun ideoiden esitysten sisältöä ei tarvitse automaattisesti analysoida, voidaan järjestelmässä välittää ideoita niiden esitystavasta riippumatta.

Jotta tietoa ideoiden väleillä vallitsevista yhteyksistä voitaisiin tehokkaasti hyödyntää, tulisi linkin merkityksestä tietää jotain. Tämä pitää paikkansa riippumatta siitä, onko linkin hyödyntäjä ihminen vai pyritäänkö tietoa linkin olemassaolosta hyödyntämään automaattisesti. Osittain linkin merkitystä voidaan päätellä siitä, mitkä resurssit se yhdistää. Pragmaattisen, tiedon käyttöä kuvaavan linkin merkitystä ei kuitenkaan voi johtaa mistään, vaan tällaisen linkin olemassaolo tuo oleellisesti järjestelmään uutta tietoa. Jotta tätä tietoa voidaan automaattisesti hyödyntää, on linkin tekijän jollain tavalla ilmoitettava linkin merkitys (esim. 'idea b *perustuu* ideaan a'). Linkkien tyypittämiseen hypermedia ei teknologiana ota kantaa, vaan käytettävät tyypit riippuvat sovellusalueen mallista. Haasteena onkin kehittää linkkityypistö, joka on sekä yleinen ja yksinkertainen että ilmaisuvoimainen.

Kontekstin kuvaamisen kannalta hypermedia on keskeinen tekniikka. Ilman linkkejä resurssin kuvaus koostuisi pelkästään resurssin omasta paikallisesta metatiedosta. Linkkien avulla muiden resurssien metatietokuvauksia voidaan hyödyntää resurssin kuvauksessa. Erityisesti tästä seuraa, että yksittäisen metatietokuvauksen ei tarvitse olla kattava, koska resurssin kuvaus muodostuu monen metatietokuvauksen yhdistelmästä. Jotta ratkaisu olisi tehokas, tulee metatietokuvausten olla sellaisia, että ne sisältävät hyödyllistä tietoa myös muiden, resurssiin linkkien kautta välillisesti liittyvien resurssien näkökulmasta. Perinteisiä metatietomalleja, kuten kirjastojen käyttämiä luokitusstandardeja, ei yleensä ole kehitetty tästä näkökulmasta. Siksi on erikseen kehitettävä sellaisia metatietomalleja, joissa hypermedian mahdollisuudet otetaan huomioon.

### 2.3 Lähdeviittausten hyödyntäminen

Sellainen hyödyllinen tieto on hyvin harvinaista, jota ei perustella muulla tiedolla tai joka ei liity muuhun tietoon mitenkään. Siten lähes jokaisessa varteenotettavassa tiedon esityksessä on viittauksia muihin esityksiin. Koska lähteiden käyttö on vakiintunut osa ideoiden välitystä, se luo hyvän peruslähtökohdan ideoiden välisten suhteiden käytön tarkasteluun.

Yksittäinen esitys sisältää lähdeluettelossaan viittaukset niihin esityksiin, joihin viittaamisen tiedon tuottaja on nähnyt tarpeelliseksi. Viitattujen esitysten lisäksi esitykseen liittyy välittömästi myös toinen esitysjoukko - nimittäin ne esitykset, joista on viittaus kyseiseen esitykseen. Jokainen viite siis yhdistäessään viittaavan esityksen viitattuun esitykseen myös luo loogisen yhteyden viitatusta esityksestä viittaavaan esitykseen.

Viittausten käytön yleisenä tiedon etsintää ja valintaa tukevana menetelmänä esitteli Garfield vuonna 1955 [Gar55]. Idea on yksinkertaisuudessaan seuraava. Mikäli löydetty esitys osoittautuu hyödylliseksi, sekä viitatut että viittaavat esitykset ovat myös hyvin potentiaalisesti hyödyllisiä. Viitatut ja viittaavat esitykset sisältävät usein tietoa, joka helpottaa tarkasteltavan tiedon ymmärtämistä ja arviointia. Erityisesti viitatut esitykset voivat sisältää relevanttia lisätietoa aiheesta ja viittaavat esitykset tietoa tarkasteltavan tiedon merkityksestä ja käyttökelpoisuudesta. Jo viittausten määrät saattavat kertoa jotain tiedon ominaisuuksista.

Perinteisten tiedonvälitysjärjestelmien kohdalla ongelmana on viittaavien esitysten löytäminen, koska joudutaan tarkastamaan kaikki myöhemmin julkaistut esitykset. Garfield ratkaisee tämän siten, että erityiset organisaatiot suorittavat tätä viittaustietojen keruuta seuraamalla jatkuvasti julkaistavia esityksiä. Käytännössä tällaisia viitetietokantoja onkin ollut tarjolla vain kolmansien osapuolien taholta, joista lähes monopoliasemassa ovat olleet Garfieldin itsensä organisoimat, Institute for Scientific Information®:in (ISI®) julkaisemat viitetietokannat, kuten Science Citation Index® (SCI®). Tällaisten viitetietokantojen käyttö on erittäin kallista ja niiden

julkaisijoilla on kaikki mahdollisuudet kontrolloida sitä, minkä tyyppistä ja kenen tuottamia ja julkaisemia tietoja niihin sisällytetään. Siksi on perusteltua kehittää avoimia tekniikoita, joiden avulla viitetietoja voivat hyödyntää nekin ideoiden tuottajat, jotka eivät saa ideoitaan viitetietokantoihin indeksoitaviksi ja nekin ideoiden tarvitsijat, joilla ei ole varaa käyttää viitetietokantojen palveluita. Ennen kaikkea automaattisen viittausten hallinnan merkitys perinteisiin menetelmiin nähden on kuitenkin se, että silloin voidaan toteuttaa palveluja, joiden toteuttaminen perinteisin menetelmin ei olisi lainkaan mahdollista.

Viittaavien esitysten muodostaman esitysjoukon hallinta on vaikeata, koska viittaavan esityksen tiedot eivät automaattisesti näy viitatussa esityksessä ja viittaavien esitysten määrä voi dynaamisesti kasvaa. Hypermediateknologia ja tietoverkon tietojenkäsittelyllinen tehokkuus antavat kuitenkin mahdollisuuden hallita ja hyödyntää viittauksia automaattisesti. Esimerkiksi ResearchIndex (CiteSeer) tarjoaa viitetietokannan ominaisuuksia automaattisen viitteiden indeksoinnin (autonomous citation indexing) avulla [LBG99, LGB99]. ResearchIndex toimii siten, että tiettyjen avainsanojen, kuten technical report, conference, proceedings, publications tms. perusteella etsitään verkosta html-sivuja, jotka sisältävät hypertekstilinkkejä joko PDF- tai PostScript-muotoisiin esityksiin. Näin löydettyistä esityksistä pyritään eristämään lähdeluettelossa olevat viittaukset sisällön analysoinnin tekniikoiden avulla. Menetelmä on luonnollisesti rajoittunut eikä sovellu yleiseksi ratkaisuksi, mutta se osoittaa, että viitetietoja ylipäättänsä voidaan automaattisesti hallita.

Toinen esimerkki viittaustietojen käytöstä on Google -hakukone, joka PageRank-algoritmin avulla pyrkii kontrolloimaan käyttäjälle esitettävää tietoa siten, että hypertekstisolmuja, joihin on paljon linkkejä, pidetään muita arvokkaampina [Pag98]. Menetelmää ei voi pitää kovin kehittyneenä, koska viittausten määrä saattaa riippua hypertekstisolmun arvon lisäksi tai jopa sen sijaan useista muistakin tekijöistä, kuten ennen kaikkea julkaisijan yleisestä tunnettuudesta [ATH00]. Silti sekin osaltaan osoittaa, että tietoverkkoon voidaan toteuttaa perinteisissä järjestelmissä mahdottomia tiedon etsinnän ja valinnan apuvälineitä. Cronin esittelee artikkelissaan useita tällaisia ratkaisuja [Cro00].

Trigg kehittää edellisiä huomattavasti sofistikoituneempaa ratkaisua ensimmäisessä hypermedia-alan väitöskirjassa vuodelta 1983 [Tri83]. Viittausten lukumääriin perustuvien bibliometrinen mittareiden yleinen heikkous on se, että resurssin ja viittausten tyyppiä ei oteta huomioon [Gar79]. Edellä esitetyt tietojenkäsittelylliset ratkaisut sisältävät saman heikkouden, minkä lisäksi kaikkien viittaavien linkkien löytämiseksi on edelleen analysoitava kaikki saatavilla olevat esitykset, mikä laajassa tiedonvälitysjärjestelmässä on melkoisen raskas operaatio. Trigg hahmottelee hypermediaperustaista tiedonvälitysjärjestelmää, jossa esitykset, linkit ja niiden tyypit merkitään siten, että ne voidaan automaattisesti tunnistaa ja niitä voidaan laskennallisesti käsitellä. Hypermediatekniikat mahdollistavat myös kaksisuuntaisten linkkien käytön, jolloin viittaavien esitysten löytäminen ja ylläpito muuttuu erittäin vaikeasta operaatiosta lähes triviaaliksi.

Sovellusten käytöstä saatujen kokemusten perusteella vaikuttaa siltä, että esitys- ja linkkityyppien hyödyntämisen suurin este ovat käyttäjät [ShM99, BuH94]. Esitysten tuottajat eivät pääsääntöisesti jaksa tai viitsi liittää kuvaustietoa esityksiin. Jos kuvaustietoa käytetään, sitä käytetään usein epäloogisesti tai ainakin niin, että kuvaustietoa hyödyntävät päättelymenetelmät eivät toimi oletetulla tavalla.

Toteutetuissa sovelluksissa esitysten ja linkkien tyypittämisen tarkoituksena on ollut lähinnä ongelmanratkaisuprosessien tehostaminen. Kontekstin kuvaamisen kannalta esitysten ja lähteiden kuvaamisen tarkoituksena ei ole tämän prosessin kuvaaminen, vaan prosessin tulosten kuvaaminen. Eron ansiosta monet prosessin kuvaamisen ongelmat ovat vältettävissä. Motivaatio resurssien kuvaamiseen lienee korkeampi, kun tiedossa on täten saavutettava resurssien näkyvyyden ja erottuvuuden lisäys. Kuvauksen luonnin ei tarvitse häiritä tuottoprosessia, kun kuvaus voidaan tehdä vasta prosessin lopuksi. Uuden idean kehittäminen voi hyvinkin kestää monia henkilötyövuosia, joten muutamassa minuutissa tapahtuva kuvaustiedon lisääminen ei tähän nähden ja siitä saatavat edut huomioon ottaen ole välttämättä ylivoimaista. Näistä eduista huolimatta motivaation ja osaamisen puutteen aiheuttamat ongelmat pysyvät kuitenkin suurimpana esteenä kuvaustiedon käytölle. Siksi on kehitettävä kompromissiratkaisuja, joissa kuvaustiedon käytön helppous ja kuvaustiedon tuoma tehokkuuden lisäys ovat sopivassa tasapainossa.

## 3 Kontekstin malli

### 3.1 Kontekstin käsite

Konteksti on yleiskäsite, jonka merkitystä on vaikea määritellä (vrt. tieto). Konteksti on tärkeä käsite mm. kieliteoriassa, tekoälyssä, psykologiassa, kognitiotieteessä ja filosofiassa, eli yleensäkin alueilla, joilla tarkastellaan merkitystä. Viime aikoina kontekstin käsitteeseen tiedon merkityksen kannalta on kiinnitetty yhä enemmän huomiota, esimerkkeinä mainittakoon pelkästään kontekstiin keskittyvät konferenssit kuten [Akm01], konteksti tekoälyn tutkimuksessa [McC87, AkS96, GhG01], kontekstin merkitys tietämyksen hallinnassa (Knowledge Management) [Abe98] ja kontekstin merkitys oppimisessa ja opetuksessa [Cob93]. Dervin esittelee joitakin näkemyksiä kontekstista lähinnä sosiaalitieteiden näkökulmasta [Der96].

Tiedon merkitys riippuu aina yhteydestä, johon tieto liittyy ja tiedon hyödyllisyys yhteydestä, jossa tietoa käytetään. Tiedon merkitys riippuu siis tiedon esityksen syntaksin ja semantiikan lisäksi myös pragmaattisista seikoista, kuten siitä, mihin tietoa on tarkoitettu käytettävän ja miten tiedon esitys tulkitaan. Tiedon välitykseen liittyvät kontrollointi- ym. sosiaaliset aspektit ovat tiedon välityksen tehokkuuden kannalta yhtä tärkeitä kuin tietoresurssien siirto lähettäjältä vastaanottajalle [BrD96]. Tiedon tallennus ja siirto eivät riitä, jos tietoa ei pystytä tulkitsemaan tai se tulkitaan väärin [BaB97]. Tietoa on erittäin vaikea esittää sellaisessa muodossa, jossa kaikki tulkitsijat ymmärtäisivät tiedon tuottajan tarkoittamalla tavalla [Her99]. Nämä näkökulmat valottavat kontekstin merkitystä tiedon välityksessä. Jotta konteksti voitaisiin ottaa huomioon myös tietojenkäsittelyllisiä ratkaisuja kehitettäessä, tulee kontekstia mallintaa muodossa, jossa sitä voidaan automaattisesti käsitellä.

Tiedon välitykseen liittyy oleellisesti kaksi kontekstin osa-alueita. Toinen on tiedon konteksti ja toinen on tiedon tarvitsijan konteksti. Tiedon arviointi ja ymmärtäminen on helppointa, kun nämä kontekstit kohtaavat, mikä on mahdollista vain omia ideoita käytettäessä. Mitä enemmän kontekstit eroavat toisistaan, sitä tärkeämpää on löytää välineitä, joiden avulla tieto saadaan välittymään tehokkaasti kontekstista toiseen.



Tiedon tarvitsijan kontekstin mallintaminen on yksi osa yleisempää käyttäjän mallintamista (user modeling, kts. [Kob93, Fis01]). Käyttäjän mallintamisella ja (idean) kontekstin mallintamisella on kuitenkin eri päämäärät. Käyttäjän mallintamisen tarkoituksena on saada kone ymmärtämään käyttäjää, kontekstin mallintamisen tarkoituksena on lisäksi saada käyttäjät ymmärtämään toisiaan. Käyttäjän mallintaminen on yleisempi menetelmä, jolle voidaan kehittää sovelluksia lähes kaikkiin käyttökohteisiin, mutta kontekstin mallintaminen on mielekästä vain, kun järjestelmässä on useita käyttäjiä, joiden 'tiedon kontekstit' eroavat toisistaan riittävästi. Eroista johtuen kontekstin osa-alueita voidaan tarkastella erillisinä aiheina. Jatkossa keskitytäänkin kontekstin mallintamiseen. Käyttäjän mallintamista sivutaan kuitenkin luvussa 5.4.

Kontekstitieto jaetaan esitettävässä mallissa kolmeen osaan. Nämä ovat **tuottokonteksti**, **julkaisukonteksti** ja **käyttökonteksti**. Jaon tarkoituksena on helpottaa idean kontekstiin liittyvien piirteiden hahmottamista. Lisäksi jako pyrkii vastaamaan myös käytännössä niitä idean välittymisprosessin vaiheita, joissa idean konteksti muodostuu.

Tuottokonteksti tarkoittaa idean tuoton aikana vallitsevaa kontekstia. Tuottokontekstin liittäminen idean metatietokuvaukseen on idean tuottajan (tuottajien) tehtävä. Tuottokontekstin kuvauksen tarkoituksena on helpottaa ideoiden löytämistä ja ymmärtämistä. Tuottokontekstia kuvaava tieto voi vastata esimerkiksi seuraaventyypisiin ideoita koskeviin kysymyksiin: kuka, milloin, miksi, kenelle, minkä tyyppinen, mistä näkökulmasta, millä tarkkuustasolla, millä oletuksilla, mihin perustuen? Koska tuottokonteksti kuvaa idean tuottoaikaista kontekstia, sitä kuvaava metatieto ei koskaan muutu.

Jokaiseen ideaan liittyy vähintään yksi esitys, jossa idea esitetään ja jokaiseen esitykseen liittyy vähintään yksi esityksen julkaisu, joka mahdollistaa esityksen löytämisen. Idean julkaisukonteksti kuvaa idean esitykseen ja julkaisuun liittyviä tietoja. Julkaisukontekstin kuvaaminen on esitysten tuottajien ja julkaisijoiden tehtävä. Julkaisukontekstia kuvaava tieto voi vastata esimerkiksi seuraaventyypisiin kysymyksiin: kenelle idea on tarkoitettu, ketkä ovat arvostaneet ideaa, millaisen arviointiprosessin idea on läpäissyt, missä muodoissa idea on esitetty ja mistä idea on

saatavilla. Julkaisukontekstin kuvaus on luonteeltaan dynaaminen, koska yksi idea voidaan esittää monella tavalla ja tietty esitys voidaan julkaista monessa eri julkaisussa.

Käyttökonteksti tarkoittaa idean käyttöä koskevaa tietoa. Idean käyttökonteksti kuvaa sitä, miten ideaa on käytetty sen julkaisemisen jälkeen. Käyttökontekstin mallintaminen mahdollistaa ideoiden etsimisen ja arvioinnin niiden käyttöä kuvaavan tiedon perusteella. Kuten julkaisukonteksti, myös idean käyttökonteksti on luonteeltaan dynaaminen. Kolmesta kontekstin osa-alueesta käyttökontekstia voi pitää arvioinnin kannalta tärkeimpänä. Käyttökonteksti on myös ollut näistä se, jonka hyödyntäminen perinteisin tiedon välityksen keinoin on ollut kaikkein hankalinta. Hyödyllisyydestään huolimatta käyttökonteksti ei tietenkään voi olla ainut idean kontekstin lähde, koska uusilla ideoilla ei ole käyttökontekstia laisinkaan, vaan pelkästään tuottokonteksti ja julkaisukonteksti.

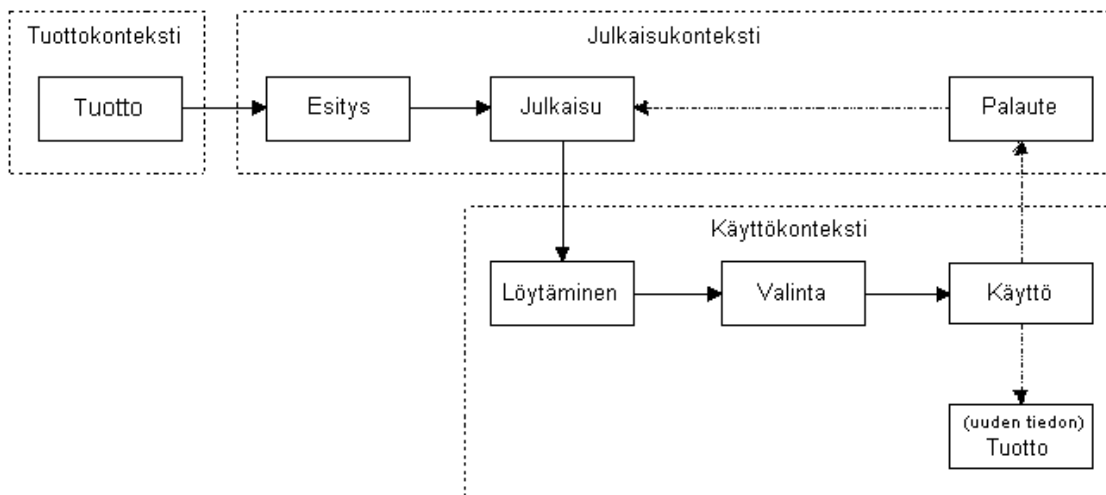
### 3.2 Ideoiden välittymisprosessi

Kontekstia ei voi kuvata, ennen kuin on olemassa käsitys mallinnettavasta kohteesta, ts. kontekstin malli. Kun ideoiden merkityksen oletetaan riippuvan myös niiden kontekstista, ei tehokasta ideoiden välitystä voi palauttaa pelkäksi informaatiopakettien siirtämiseksi lähettäjältä vastaanottajalle. Koska tarkoituksena on kehittää uusia ideoiden välittymistä tukevia menetelmiä, joiden käyttö perinteisessä tiedon välityksessä ei ole ollut mahdollista, ei mallia voi ottaa suoraan perinteisistä järjestelmistä. Mallien kehittäminen ja ymmärtäminen on vaikeaa. Siksi tässä pysytelläänkin hyvin yleisillä linjoilla.

Vain sellaisilla ideoilla on merkitystä, joita käytetään jotenkin. Jotta oikeita ideoita käytettäisiin oikeissa tilanteissa, ideoita pitää pystyä valikoimaan onnistuneesti. Hyödyllisen tiedon olemassaolo sinänsä ei riitä, jos käytettävä tieto valitaan sattumanvaraisesti. Tästä näkökulmasta toimivien tiedon valinnan ja arvioinnin menetelmien tarjoaminen on tiedonvälitysjärjestelmän tärkein ja kriittisin ominaisuus. Tehokkaat valintakriteerit johtavat myös siihen, että uudet ideat voivat perustua

parhaisiin jo olemassa oleviin ideoihin, jolloin ideoiden kehittyminen voi olla progressiivista. Tiedon valinnan merkitystä tiedon kehittymiselle ei tämän tutkielman puitteissa ole mahdollista tarkemmin perustella, aihetta käsitellään syvällisemmin muun muassa Popperin ja Hullin artikkelikokoelmissa [Pop72, Hul01]. Perusnäkömyksenä on, että tiedon arvon lopulliseksi määrittämiseksi ei ole vedenpitäviä menetelmiä, vaan päin vastoin voidaan puhua jopa yritys- ja erehdys – menetelmistä. Siksi on tärkeää, että tiedon tarvitsijalla itsellään on mahdollisimman monipuoliset keinot, vastuu ja valta ideoiden arvon määrittämisessä ja että myös tieto tiedon arvoa määrittävistä kriteereistä voi kehittyä kuten muukin tieto.

Ideoiden välittymisen malli koostuu prosessista, joka jakaantuu kahdeksaan vaiheeseen. Tuottokonteksti muodostuu tiedon tuottovaiheessa. Julkaisukonteksti muodostuu esitys-, julkaisu- ja palautevaiheissa. Käyttökontekstiin kuuluu neljä vaihetta, jotka ovat löytäminen, valinta, käyttö ja uuden tiedon tuotto. Idean käyttö uuden idean tuottamisessa voi siten kuulua sekä tuotto- että käyttövaiheeseen riippuen siitä, kumman idean kannalta asiaa tarkastellaan. Ideoiden välittymisen malli on esitetty kuvassa 1. Nuolet kuvaavat tiedon siirtymistä prosessin vaiheista toiseen.



Kuva 1. Ideoiden välittymisprosessi

*Tuotto.* Uusien ideoiden tuottaminen on usein luonteeltaan pitkä, vuosiakin kestävä luova ja 'sumeä' prosessi. Siksi tuottoprosessia ei kannata kuvata suoraan, vaan on keskityttävä sellaisiin yleisiin tuottoa kuvaaviin tietoihin, jotka tuottaja voi erikseen liittää valmiiseen ideaan. Tärkein idean tuottokontekstiin kuuluva tekijä, jota järjestelmässä voidaan hyödyntää, ovat ne ideat, joita uuden idean tuottamisessa käytetään hyväksi eli idean lähteet.

*Esitys.* Kun uusi idea on luotu, on se esitettävä, jotta muutkin kuin idean tuottaja(t) voisivat hyödyntää sitä. Määritellään idean esitys idean esittämisestä tehdyksi tallenteeksi. Idea voidaan esittää esimerkiksi kirjoittamalla, jolloin tuloksena syntyvä teksti on yksi idean esitys. Idea voidaan esittää myös vaikkapa pitämällä esitelmä. Esitelmää koskevat käsikirjoitus, referaatti, esitelmäpuheen nauhoitus ja esitelmästä kuvattu video ovat kaikki eri esityksiä. Toisaalta yhden esityksen tallentaminen eri muodoissa, kuten LP-, CD-, tai DVD-levylle, ei tuota uusia esityksiä, vaan kaikki ne sisältävät saman esityksen. Tämä pätee tietenkin sillä varauksella, että tallennusmuodon muuttaminen ei merkittävästi muuta esityksen tietosisältöä. Yhdestä ideasta voidaan luonnollisesti tuottaa useita eri esityksiä. Samoin yksi esitys voi sisältää monta ideaa, mutta aina vähintään yhden. Esityksellä voi olla eri tuottajia kuin itse idealla. Esimerkiksi henkilö, joka kääntää esityksen toiselle kielelle on esityksen tuottaja, mutta ei välttämättä itse idean tuottaja.

*Julkaisu.* Idean julkaisu tarkoittaa idean esittävän esityksen tallentamista ja saataville asettamista johonkin tietovarastoon. Jokainen esityksen julkaisu tuottaa uuden dokumentin. Siten esimerkiksi esityksen julkaiseminen toisessa tietovarastossa ja esityksen julkaiseminen toisessa tallennusmuodossa tuottavat aina uuden dokumentin. Yhtä julkaistua resurssia kutsutaan dokumentiksi. Yleisimmin jokainen esitys julkaistaneen erikseen, mutta dokumentti voi olla myös useamman esityksen kokoelma. Esimerkki tällaisesta dokumentista on artikkelikokoelma, jossa ideoiden esityksiä on koottu yhteen. Kaiken kaikkiaan yksi dokumentti voi siis sisältää huomattavan monta ideaa ja yhdestä ideasta voi olla saatavilla huomattavan monta dokumenttia. Tietovaraston hallinnoijaa (ylläpitäjää) kutsutaan julkaisijaksi. Julkaisijan tehtävä on valikoida ja arvioida ideoita tiedon tarvitsijoiden puolesta, tarjota kanava, joka helpottaa saman aiheisten ideoiden löytämistä sekä toimia idean luotettavuuden, alkuperäisyyden ja saatavuuden varmentajana.

*Löytäminen.* Ennen kuin tietoa voidaan käyttää, se on löydettävä. Tietoa löydetään etsimällä. Onnistuneen etsimisen lopputuloksena on yhden tai useamman sellaisen idean löytyminen, jotka mahdollisimman todennäköisesti vastaavat tiedon tarvetta.

*Valinta.* Onnistuneen etsinnän tuloksena on joukko löydettyjä ideoita. Löydetty ideat on arvioitava ja niistä valittava osa käyttöä varten. Mitä enemmän potentiaalisesti hyödyllisiä ideoita on, sitä enemmän aikaa ideoiden valinta vie. Valinnan toimivuus on siinä mielessä kriittistä, että hyödyllisten ideoiden löytymisestä ei sinänsä ole hyötyä, mikäli hyödyllisiä ideoita ei myös pystytä erottamaan kaikkien löydettyjen ideoiden joukosta.

Tiedon etsintää ja arviointia tehdään useimmiten iteratiivisesti, eli välillä tarkastellaan löydettyjä ideoita ja välillä niitä haetaan lisää. Siksi löytämisen ja valinnan erottaminen toisistaan ei käytännössä ole useinkaan helppoa. Teknisesti löytäminen ja valinta voidaan kuitenkin erottaa seuraavasti. Löytäminen tarkoittaa dokumenttien tai niiden kuvausten hakemista ja valinta haettujen dokumenttien tai niiden kuvausten tarkastelua. Mitä enemmän löydettyjä dokumentteja on, mitä vaikeampi dokumentteja on hankkia ja mitä vaikeampaa dokumenttien sisällön ymmärtäminen on, sitä enemmän valinnassa joudutaan turvautumaan dokumentin sisällön sijaan saatavilla olevaan dokumentin kuvaukseen. Juuri valintaa helpottavan kuvauksen luomisessa kontekstitiedon olemassaolosta on kaikkein eniten hyötyä.

*Käyttö.* Ideaa voidaan käyttää lukemattomilla eri tavoilla. Vain pieni osa käyttöä koskevasta tiedosta saadaan jollain tavalla kerättyä ja tallennettua. Mallissa lähdetään siitä, että käyttäjiltä voidaan saada käyttöä koskevaa tietoa kahdella tavalla – tietoa koskevana palautteena (kritiikkinä) sekä silloin, kun ideaa käytetään uuden idean tuotossa.

*Palautte.* Idean arviointia itsessään voidaan pitää yhtenä idean käyttökohteena. Arviointityön ja –tiedon jakaminen muiden tiedon tarvitsijoiden kanssa on yksi keino helpottaa arviointia. Jos palautteen halutaan olevan yleisesti käytettävissä, se voidaan julkaista uutena ideana tiedonvälitysjärjestelmässä. Tällaisen palautteenantotavan

lisäksi myös yksinkertaisemmat ja nopeammat, viestintään perustuvat palautekanavat voivat olla hyödyllisiä.

Erilaisten palautteenantotapojen kirjo on rajaton. Ensimmäisiä tällaisten palvelujen hahmotteluja sisältää Swansonin artikkeli vuodelta 1964 [Swa64], joista mainittakoon käyttäjien arvioiden, muistiinpanojen ja huomautusten (annotations) liittäminen ideoihin. Twidale ja Nichols esittelevät erilaisia ratkaisuja erityisesti tietokoneavusteisen yhteistyön näkökulmasta [TwN98]. Gresham korostaa virtuaalisten konferenssien ja keskusteluryhmien merkitystä viestintäkanavana [Gre94]. Yksi esimerkki palautekanavan käytöstä on, että sen avulla idean tuottaja voisi oikaista idean ymmärtämiseen liittyviä väärinkäsityksiä ja esittää selventäviä esimerkkejä. Schatz esittelee ansiokkaasti informaalin, viesteihin perustuvan tiedon välityksen menetelmiä osana formaalimpaa, dokumentteihin perustuvaa tiedonvälitysjärjestelmää [Sch92].

Kuten dokumentitkin, myös palaute on asetettava saataville. Luonnollisimmin palaute voidaan tallentaa samaan tietovarastoon, johon palautteen kohteena oleva dokumentti on tallennettu. Kuvassa 1 palautteen kulku on merkitty katkoviivalla merkityllä nuolella sen korostamiseksi, että idean sijaan julkaistaan ideaa koskevia arvioita. Palautteen hallintamekanismeja ei jatkossa käsitellä tämä tarkemmin. Voidaan olettaa, että dokumentin metatietokuvaukseen liitetään viestintäkanavien tunnisteet, jolloin palautteeseen päästään metatietokuvauksen kautta käsiksi, mutta sitä hallitaan metatietokuvauksista erillisenä ja julkaisijat voivat tarjota erilaisia viestintäpalveluja toisistaan riippumattomasti.

*Käyttö uuden idean tuotossa.* Kontekstin kuvaamisen kannalta oleellista on sellainen idean käyttö, josta seuraa uuden idean julkaiseminen. Tämä toteutuu, kun ideaa käytetään uuden idean tuottamisen apuna tai kun uusi idea tarkastelee vanhaa ideaa, esimerkiksi raportoi idean merkityksestä käytännössä. Jo sen tietäminen, kuinka paljon ja ketkä tiettyä ideaa ovat käyttäneet, voi olla hyödyllistä. Vähintään yhtä hyödyllistä on tietää myös, mihin ideaa on käytetty ja millaisia saadut kokemukset ovat olleet. Kaikki tällaiset tiedot voidaan julkaista uusina ideoina, jotka muodostavat osan lähteenä tai kohteena olevan viitatus idean käyttökontekstia.

Ideoiden välittymisen malli tuottaa siis seuraavat resurssityypit ja niiden väliset suhteet. Tuottajat tuottavat ideoita ja esityksiä. Yksi esitys voi sisältää monta ideaa ja yksi idea voidaan esittää monessa esityksessä. Julkaisijat julkaisevat esityksiä dokumentteina, jotka ovat saatavilla julkaisijan ylläpitämän tietovaraston kautta. Yksi esitys voidaan julkaista monessa dokumentissa ja yksi dokumentti voi sisältää monta esitystä. Tietovaraston kautta voidaan julkaista myös ideoita koskevaa palautetta. Malliin perustuen idean kontekstia kuvaava tieto on siis tietoa idean suhteista muihin ideoihin ja palautteeseen, idean ja esitysten tuottajiin, esityksiin, dokumentteihin sekä julkaisijoihin.

Mallin luomisessa pääperiaatteina ovat olleet yksinkertaisuus, jotta kontekstitiedon luominen muodostuisi mahdollisimman helpoksi, sekä yleisyys, jotta mallin käyttäjäkunta olisi mahdollisimman laaja. Vastaavalla tavalla kehitettyjä ja vastaaviin periaatteisiin perustuvia idean kontekstin malleja ei ole. Mallia voidaan kuitenkin verrata International Federation of Library Associations:in (IFLA) kehittämään yleiseen malliin bibliografisten kuvausten pohjaksi [Pla98]. Huomionarvoisimmat erot ovat ero mallin johtamistavassa sekä mallin kattavuus. Tässä malli perustuu tiedon välittymisen hahmottamiseen prosessina, kun IFLA:n malli on staattinen kuvaus, joka pitkälti perustuu jo olemassaoleviin kuvauksiin. IFLA:n mallissa on pyritty kattavuuteen, eli kuvaamaan kaikentyyppisiä teoksia, kun tässä on keskitytty yhteen teosten osajoukkoon, ideoihin. IFLA:n malli on siten sekä monipuolisempi että monimutkaisempi kuin tässä esitetty.

Erojen sijaan tärkeintä on havaita mallien samankaltaisuus. Tässä esitetty malli voidaan tulkita IFLA:n mallin kanssa yhteensopivaksi, kun resurssityypeille (IFLA:ssa kuvauksen kohde on *entity*) asetetaan seuraava vastaavuus: idea = concept, esitys = expression ja dokumentti = manifestation. Sitä, että kahdella toisistaan riippumattomalla tavalla päädytään yhteensopiviin malleihin, voitaneen pitää ainakin jonkinlaisena merkinä mallien yleisluontoisuudesta.

### 3.3 Avoin arkkitehtuuri

Jotta tietoa voitaisiin siirtää paikasta toiseen, tarvitaan tiedonsiirtokanavia. Jotta tietoa voitaisiin siirtää ajasta toiseen, tarvitaan tietovarastoja. Tietovarastot ja tiedonsiirtokanavat luovat infrastruktuureja, joihin tiedonvälitysjärjestelmien käytännön toteutus perustuu. Tiedonvälitysjärjestelmän arkkitehtuuri kuvaa sitä, miten tiedonsiirtokanavia ja tietovarastoja käytetään.

Arkkitehtuureja voidaan kehittää niiden teknisten ominaisuuksien parantamiseksi. Tallennuskapasiteetin ja tiedonsiirtonopeuden lisäksi tärkeitä aspekteja ovat tällöin esimerkiksi vikasietoisuus, skaalautuvuus ja hajautuvuus. Tässä arkkitehtuuria ei kuitenkaan tarkastella teknisten ominaisuuksien parantamiseksi. Siksi ajatellaan, että nykyiset teknologiat ovat riittävän tehokkaita. Toteutuskysymyksistä kiinnostuneiden kannattaa tutustua esimerkiksi määritelmiin Hypertext Transfer Protocol (HTTP) [Fie99], Common Object Request Broker Architecture (CORBA) [Obj02], Lightweight Directory Access Protocol (LDAP) [WHK97] sekä ns. Peer-to-Peer –ratkaisujen edustajista määritelmään EDUTELLA [Nej02]. Arkkitehtuuria hahmotellaan tässä sellaisella tasolla, että järjestelmän toteuttamiseksi ei tarvita uusia tietovarastoja tai tiedonsiirtomenetelmiä koskevia innovaatioita.

Toteutuksen tehokkuuden sijaan arkkitehtuurin merkitys nähdään siinä, että se määrittää tiedonvälitysjärjestelmän avoimuuden perustason. Internet tiedonsiirtokanavana on luonteeltaan avoin. Kuka tahansa voi kytkeytyä Internetiin ja periaatteessa olla yhteydessä mihin tahansa muuhun Internetiin liitettyyn tietovarastoon. Samoin on olemassa useita määritelmiä, kuten kaikki edellä mainitut, jotka ovat siinä mielessä avoimia, että niiden käyttöä ei rajoiteta, niiden määritelmät ovat vapaasti saatavilla ja määritelmät eivät ole laitteisto/käyttäjärjestelmäriippuvia. Teknisen ja toteutuksen avoimuuden sijaan tässä tarkastellaankin erityisesti käytön avoimuutta.

Käytön avoimuus voi olla monentasoista. Julkaiseminen voi olla vapaata tai säädeltyä, julkaisijana toimiminen voi olla vapaata tai säädeltyä ja julkaisujen käyttö voi olla vapaata tai säädeltyä. Yhteensä täysin suljetusta täysin avoimeen järjestelmään on siis



kahdeksan erilaista avoimuusluokkaa. Teknisesti yksinkertaisin arkkitehtuuri on sellainen, jossa on vain yksi tietovarasto. Tällainen ratkaisu ei kuitenkaan ole avoin, koska tietovaraston ylläpitäjä voi täysin kontrolloida sitä, ketkä voivat käyttää järjestelmää ja millaisia palveluja järjestelmä tarjoaa.

Toinen esimerkki teknisesti avoimesta, mutta käytön suhteen osittain suljetusta arkkitehtuurista on Open Archives Initiative (OAI) [LaV01, Lag02]. Siinä jokaisessa tietovarastossa hallitaan vain tietovaraston omia tietoja, mutta tietoja kuvaava metatieto kootaan keskitetysti yhteen (meta)tietovarastoon, joka voi toki olla replikoituna useaan paikkaan. Malli sallii useamman tietovaraston perustamisen, mutta metatietokannan ylläpitäjä päättää, minkä tietovarastojen metatietoa metatietovarastoon kerätään ja millaista toiminnallisuutta metatietovarasto ja siten järjestelmä kokonaisuudessaan sisältää. Näistä esimerkeistä poiketen tässä esitellään arkkitehtuuri, joka on käytön suhteen avoin. Käytännössä tämä tarkoittaa sitä, että kaikki tietovarastot ovat samanarvoisessa asemassa ja ne voivat tarjota palveluja toisistaan riippumatta.

Arkkitehtuurin mallin perustana on informaatioavaruuden (information space) käsite. Informaatioavaruus sisältää kaiken järjestelmään mallinnetun informaation. Siten informaatioavaruuden ulkopuolella olevaa informaatiota ei mallin näkökulmasta ole olemassa. Frisse ja Cousins osittavat informaatioavaruuden kahteen osajoukkoon, jotka ovat dokumenttiavaruus (document space) ja indeksiavaruus (index space) [FrC89]. Näiden avaruuksien ero määritellään siten, että dokumenttiavaruuden informaatiota hyödyntävät menetelmät, kuten dokumenttien selaus, ovat paikallisia (local) eli ne käyttävät kerrallaan vain yhtä dokumenttia koskevaa informaatiota. Indeksiavaruuden menetelmät ovat globaaleja (global) eli niiden avulla voidaan menetelmiä kohdistaa kerralla laajempaan dokumenttijoukkoon. Käytännössä paikallisuus tarkoittaa sitä, että menetelmät perustuvat dokumentteihin itseensä ja globaalius sitä, että menetelmät perustuvat dokumenteista erilliseen tiedon kuvaukseen. Tässä informaatioavaruus ositetaan vastaavasti, mutta indeksiavaruuden sijaan puhutaan metatietoavaruudesta sen korostamiseksi, että metatietoavaruus voi sisältää tietoa, jota ei voi johtaa dokumenttiavaruuden tiedosta ja metatietoavaruus voi sisältää tietoa myös muista resursseista kuin dokumenteista, kuten ideoista, tuottajista ja julkaisijoista.

Dokumenttiavaruus sisältää järjestelmässä välitettävät ideat ja metatietoavaruus näiden ideoiden kontekstia kuvaavan tiedon. Informaatioavaruuden osittaminen metatietoavaruuteen ja dokumenttiavaruuteen luo mahdollisuuden jakaa tiedonvälitysjärjestelmä vastaavasti metatiedonvälitysjärjestelmään ja dokumenttienvälitysjärjestelmiin.

Dokumenttienvälitysjärjestelmien tehtävänä on dokumenttiavaruuteen kuuluvan tiedon hallinta, eli dokumenttien käsittely, tallennus ja siirto. Kun jokaisella dokumenttiavaruuden dokumentilla on yksikäsitteinen ja globaali tunniste, voidaan metatietoavaruudesta viitata dokumenttiavaruuden dokumentteihin ja ainakin digitaalisessa muodossa olevat dokumentit saadaan suoraan haettua (varattua, tilattua, lainattua, ostettua...) tämän tunnisteiden perusteella.

Arkkitehtuuri ei aseta erityisiä rajoituksia dokumenttienvälitysjärjestelmille, joten voidaan olettaa, että soveltuvat järjestelmät ovat käytettävissä. Erityisesti dokumenttienvälitysjärjestelmien ei tarvitse olla täysin avoimia, vaan julkaisijat voivat esimerkiksi periä maksua dokumenttien käytöstä. Jatkossa keskitytäänkin tiedonvälitysjärjestelmän mallintamisen osalta pelkästään metatietoavaruuden ja metatiedonvälitysjärjestelmän mallintamiseen.

Metatiedonvälitysjärjestelmän tehtävänä on metatietoavaruuteen kuuluvan metatiedon käsittely, tallennus ja siirto. Jokaista dokumenttiavaruuden dokumenttia vastaa metatietoavaruudessa yksi metatietotietue. Metatietotietueessa metatieto esitetään attribuuttien luettelona. Dokumentin ja sitä vastaavan metatietotietueen välinen kytkentä on olennainen, sillä se konkretisoi ideoiden kontekstin metatietotietueisiin tallennetuksi tiedoksi, jota voidaan tietojenkäsittelyllisesti hallita ja käsitellä metatiedonvälitysjärjestelmässä.

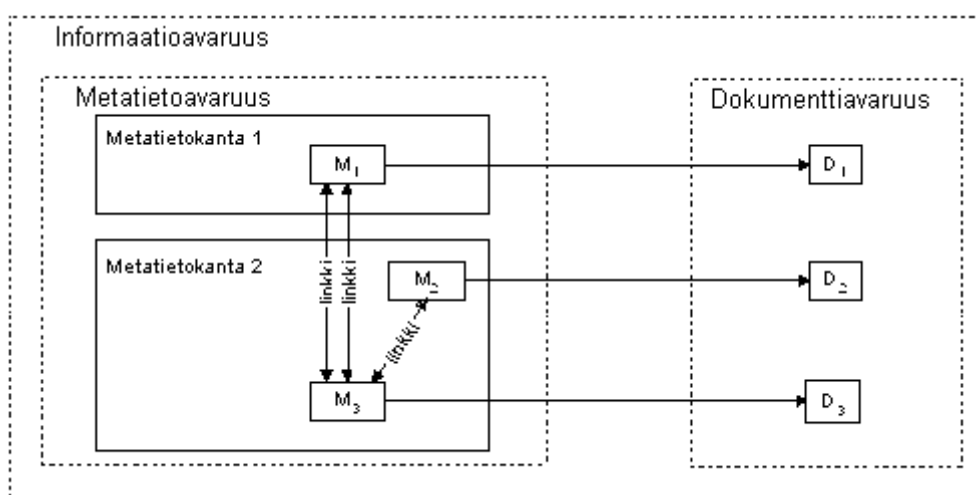
Metatietotietueesta voidaan viitata dokumentin lisäksi muihin metatietotietueisiin. Näitä viittauksia kutsutaan linkeiksi. Linkit ovat aina kaksisuuntaisia, eli viittaaminen johonkin metatietotietueeseen aiheuttaa aina viittauksen myös viitattavasta viittaavaan. Jotta metatietotietueisiin voitaisiin viitata, niillä tulee olla yksikäsitteinen ja globaali tunniste. Siihen, miten tämä tunniste muodostetaan, ei oteta kantaa. Esimerkki

potentiaalisesti käyttökelpoisesta tunnistetyypistä on Universal Resource Identifier (URI) [BFM98, Coa01].

Metatietoavaruus osittuu metatietovarastoihin. Näitä metatietovarastoja kutsutaan metatietokannoiksi niiden tietojenkäsittelyllisen luonteen korostamiseksi. Metatietovarastoja voitaisiin kutsua myös linkkikannoiksi. Resursseista riippumattomat linkkipalvelut ja linkkikannat (link services, linkbases) toteuttavat resurssien sisäisen ja resurssien välisen hypermediatoiminnallisuuden (hypertext functionality) avoimissa hypermediajärjestelmissä (open hypermedia systems). Yksi ensimmäisistä ratkaisuista on Sun's Link Service [Pea89], jossa käsite avoin hypermediajärjestelmä esitettiin. Carr, Hall, ja De Roure esittelevät myöhemmin kehitettyjä järjestelmiä [CHR99] ja yksityiskohtainen yhden toteutetun järjestelmän eli Hyperwaven kuvaus löytyy Maurerin kirjasta [Mau96].

Avoimen hypermedian idea on siinä, että tiedon välisiä suhteita käsitellään itse tiedosta riippumatta, jolloin ratkaisut eivät ole sidottuja tiettyihin tiedon esitystapoihin, ja linkkien, etenkin kaksi- tai useampisuuntaisten, hallinta yksinkertaistuu. Esitettävä arkkitehtuuri on olennaisesti avoimen hypermediajärjestelmän arkkitehtuuri, missä resursseja ovat dokumentit, metatietokannat toimivat linkkikantoina ja dokumenttien välisiä suhteita hallitaan metatietotietueiden välisten linkkien avulla.

Kuva 2 esittelee arkkitehtuuriin liittyvät käsitteet esimerkin avulla. Informaatioavaruus osittuu metatietoavaruuteen ja dokumenttiavaruuteen. Metatietoavaruuden näkökulmasta dokumenttiavaruus on dokumenttien tunnisteen joukko. Kuvassa on kolme tällaista tunnistetta,  $D_1$ ,  $D_2$  ja  $D_3$ . Jokaiseen dokumenttiin  $D_x$  liittyy metatietoavaruudessa täsmälleen yksi metatietotietue  $M_x$ . Jokainen metatietotietue on tallennettu yhteen metatietokantaan, joita kuvassa on kaksi kappaletta. Jokaisesta metatietotietueesta tiedetään, missä metatietokannassa se sijaitsee. Dokumenttien sisältämien tietojen väleillä voi olla suhteita, jotka kuvataan metatietotietueiden välisinä kaksisuuntaisina linkkeinä. Linkin tiedot on tallennettu linkin kohteina oleviin metatietotietueisiin. Kahden metatietotietueen välillä voi olla useampi kuin yksi linkki, kuten kuvassa metatietotietueiden  $M_1$  ja  $M_3$  välillä.



Kuva 2. Informaatioavaruus

Kahteen eri dokumenttiin voi liittyä yhteistä idean kontekstia kuvaavaa metatietoa, kun ne esittävät samoja ideoita. Jotta tiettyyn dokumenttiin liittyvän metatiedon saatavuus ei riippuisi muiden dokumenttien metatietotietueiden saatavuudesta tai muiden tietovarastojen toimivuudesta, on kaikki tietyn dokumentin kontekstiin suoraan kuuluva metatieto tallennettava dokumentin omaan metatietotietueeseen. Sama metatieto voi siis sijaita useassa paikassa. Ongelmana, varsinkin hajautetussa arkkitehtuurissa, on huolehtia tällöin siitä, että metatietoavaruus on sisäisesti konsistentti eli ristiriidaton.

Ristiriidattomuus voidaan varmistaa siten, että jokaisesta metatietotietueesta on aina linkit kaikkiin niihin metatietotietueisiin, joiden kanssa sillä on yhteistä metatietoa. Kun yhden metatietotietueen tiedot muuttuvat, voidaan linkkien avulla muutos suorittaa automaattisesti muihin metatietotietueisiin. Kun virhetilanteiden ja poikkeustilanteiden korjaamista vaatimia metatietotietueiden muutoksia ja poistoja ei oteta huomioon, voi metatietotietue muuttua sen julkaisemisen jälkeen vain kahdessa tapauksessa. Toinen tapaus on se, että julkaistaan dokumentti, jossa metatietokuvauksen kuvaamaa ideaa käytetään, ts. kyseisen idean käyttökonteksti muuttuu. Toinen tapaus on se, että julkaistaan dokumentti, joka sisältää samoja ideoita tai esityksiä, ts. idean julkaisukonteksti muuttuu. Toisin sanoen **metatietoavaruuden tila muuttuu ainoastaan silloin, kun julkaistaan uusi dokumentti.**

Tiedonvälitysjärjestelmän toiminnan suhteen ei ole kriittistä, että metatietoavaruuden tila olisi konsistentti joka hetki, vaan muutokset voidaan suorittaa tarvittaessa tai eräajona riittävän usein.

## 4 Kontekstin kuvaaminen

### 4.1 Idean kuvaaminen

Luvussa kolme määriteltiin idean kontekstin malli resurssityyppineen ja niiden välisine suhteineen sekä esitettiin tapa hallita kontekstietoa metatietotietueiden avulla avoimessa järjestelmässä. Tässä luvussa käsitellään tarkemmin metatiedon sisältöä eli resurssien kuvaamista metatietotietueiden attribuuttien avulla. Tarkoituksena on osoittaa, että kontekstia voidaan kuvata yksinkertaisella tavalla siten, että kontekstiedon luominen ei aiheuta järjestelmän käyttäjille, ts. ideoiden ja esitysten tuottajille ja julkaisijoille ylivoimaista rasitetta. Tutkielman tarkoituksena on myös osoittaa, että kontekstia voidaan kuvata siten, että sen avulla ja tietojenkäsittelyllisiä menetelmiä hyödyntäen voidaan toteuttaa tiedon tarvitsijan näkökulmasta tehokkaita tiedon etsinnän ja valinnan apuvälineitä. Viime mainittuun päämäärään pyritään siten, että tässä luvussa määriteltävää kontekstin kuvausta käytetään perustana luvussa viisi, jossa mahdollisuuksia tiedon löytämisen ja valinnan tukemiseen käsitellään.

Metatietotietueen attribuutit käsitellään seuraavasti. Tässä käsitellään idean tuottokontekstin kuvaamiseen liittyvät attribuutit muun kuin ideoiden välisten suhteiden osalta, luvussa 4.2 käsitellään ideoiden väliset suhteet eli tuottokonteksti siltä osin sekä käyttökonteksti ja luvussa 4.3 käsitellään julkaisukontekstin kuvaaminen eli esityksiin, julkaisijoihin ja dokumentteihin liittyvät attribuutit. Luku 4.4 on yhteenveto, jossa metatietotietueen attribuutit esitellään kootusti.

Jotta ideaan ja ideakohtaiseen metatietoon voitaisiin järjestelmän ohjelmallisella (l. toiminnallisella, automaattisella) tasolla viitata, tulee idealla olla yksikäsitteinen tunniste. Tämän attribuutin nimi olkoon **idean tunniste**.

Idean tunniste saattaa olla hyvinkin kryptinen ja semanttisesti merkityksetön. Jotta myös ihmisten olisi helppo erottaa ideat toisistaan ja keskustella ideoista, on idealla oltava myös tällaiseen käyttöön soveltuva tunniste. Tämän attribuutin nimi olkoon

**idean nimi.** Toisin kuin tunniste, nimi voi olla mikä tahansa vapaasti valittavissa oleva merkkijono.

Jotta tuottajien oikeudet ja vastuut ideoihin olisivat todennettavissa, tulee ideoiden tuottajat luetella. Tämän attribuutin nimi olkoon **idean tuottajat**. Toinen tapa ilmoittaa idean tuottajat on viitata metatietotietueessa idean tuottajien tietovarastoihin. Tätä jälkimmäistä tapaa käsitellään luvussa 4.3. Tässä yhteydessä on kuitenkin syytä korostaa sitä, että idean tuottajat ovat aina henkilöitä. Esimerkiksi organisaatioiden mahdolliset oikeudet ja vastuut ideaan ilmaistaan siten, että organisaatio julkaisee idean (kts. luku 4.3).

Se, miten ideaa on arvioitava ja idean esitystä tulkittava, riippuu siitä, milloin idea on julkaistu. Idean tuotto-, julkaisu-, ja käyttökäytännöt, kuten viittauskäytännöt, saattavat eri aikoina vaihdella. Samaten eri aikoina saatetaan käyttää erilaisia esitystyyliä, kuten erilaista argumentointia, retoriikkaa tai termistöjä. Suurin merkitys ajalla idean kontekstin suhteen on siinä, että idean käyttökäytännön määrä riippuu suoraan siitä, kuinka kauan idea on ollut käytettävissä. Näistä syistä johtuen ideoihin on merkittävä niiden tuottoajankohta. Tämän attribuutin nimi olkoon **idean tuottoajankohta**.

Kun ideoiden väleillä vallitsevia suhteita tunnetaan, idean kuvausta voidaan käyttää toisten ideoiden kontekstin kuvauksen lähteenä. Jotta kontekstin johtaminen olisi tehokasta, tarvitaan tähän tarkoitukseen soveltuvia attribuutteja. Tarkoitukseen soveltuvia kuvaustapoja on olemassa melko vähän. Suurin osa erilaisista kuvaustavoista keskittyy aiheen kuvaamiseen, mutta koska ideoilla luonnostaan on eniten suhteita muihin samansisältöisiin ideoihin, tästä tiedosta ei voi johtaa kovin merkittäviä päätelmiä.

Kun kuvauksen tuottaja on idean tuottaja, voidaan sisällöstä johdettavan tiedon lisäksi kuvata kaikkea sitä, mistä idean tuottajalla on tietoa. Tämä laajentaa huomattavasti kuvauksen mahdollisuuksia. Toisaalta, kun kuvauksen tuottaja on idean tuottaja, kuvaamisen tulee olla hyvin helppoa. Siten kovin monimutkaisia tai usein muuttamista vaativia kuvausmenetelmiä ei voi käyttää.

Edellä mainitun valossa idean tarkoituksen kuvaaminen vaikuttaa erityisen tehokkaalta. Erilaisia tarkoitukseluokkia tuskin on rajattomasti, uusia tuskin kehitetään kovin nopeassa tahdissa ja jokaisen idean tuottajan voi olettaa tietävän jotain siitä, mihin tarkoitukseen idea on tuotettu. Ennen kaikkea idean tarkoitus on hyödyllinen tieto ideaa koskevia päätelmiä tehtäessä. Yleensä tiedon tarvitsija tarvitsee tietoa johonkin tarkoitukseen. Tarkoituksen avulla nähdään helposti, millaisia lähteenä käytetyt ideat ovat olleet ja mihin tarkoituksiin ideaa on käytetty. On myös selvää, että eri tarkoituksiin tarkoitettuja ideoita tulee arvioida eri tavoin.

Tiedon tarkoituksen luokitteluja on tehty mitättömän vähän. Yhtä harvoista tähän ajatukseen perustuvista luokituksista on kehitetty ScholOnto –projektin yhteydessä [BMD99, BMD00]. Siinäkin on tosin keskitytty enemmän luokituksen tarpeellisuuden perustelemiseen kuin itse luokituksen kehittämiseen. Valmiiden luokitusten puutteesta johtuen tässä esitellään aivan uusi luokitus. Luokitusta ei voi pitää lopullisena, koska sen toimivuutta käytännössä ei ole ollut mahdollista testata.

Olkoon idean tarkoitusta kuvaavan attribuutin nimi **idean tarkoitus**. Attribuutin arvo voidaan valita kahdeksasta tarkoitukseluokasta, jotka luetellaan seuraavassa.

*Miksi-idea.* Miksi-idean tarkoitus on kuvata sitä, miksi tiettyntyyppisiä ideoita tulisi tuottaa. Se on siis eräänlainen 'meta-idea' eli idea siitä, millaisia muita ideoita tarvitaan. Idean tuottamisen tarve voi olla seurausta yleisestä asiantilasta tai ongelmasta, jolloin miksi-idea voi olla esimerkiksi poliittinen, filosofinen tai yhteiskunnallinen visio, näkemys tai spekulatio. Idean tuottamisen tarve voi syntyä myös erityisestä syystä, kuten idean tuottajien, julkaisijoiden tai jonkin organisaation tarpeesta tai mahdollisuudesta. Tällöin miksi-idea voi olla esimerkiksi strateginen linjanveto, projektisuunnitelma tai tutkimussuunnitelma.

*Mitä-idea.* Mitä-idean tarkoitus on määritellä ongelma siten, että siihen voidaan etsiä ratkaisuja tai ilmiö siten, että sen käyttäytymistä voidaan ymmärtää. Idea kiinnittää näkökulman, perusoletukset ja määritelmät, joiden avulla jotakin aihetta voidaan käsitellä. Mitä-idea on käytännössä esimerkiksi malli tai teoria.



*Miten-idea.* Idean tarkoitus on kertoa, miten jokin ongelma kannattaa tai voidaan ratkaista. Miten-idea on käytännössä esimerkiksi metodi, toimintaohje tai prosessin toiminnallinen kuvaus.

*Tulos.* Tulos-tyyppisen idean tarkoitus on esitellä yksi spesifi ratkaisu, joka on käytännössä toteutettu tai jota on käytetty. Tarkoituksena ei ole kuvata tuloksen käyttöä tai käytön seurauksia vaan pelkästään itse tulosta. Toteutus on käytännössä esimerkiksi yhden koneen, laitteen tai muun rakennelman kuvaus tai jonkin metodin tai prosessin toteutuksen kuvaus. Myös luonnontieteelliset, matemaattiset ja vastaavat johdetut vakiot kuuluvat tähän luokkaan.

*Analyysi.* Analyysin tarkoituksena on testata tai tarkastella jonkin toisen idean ominaisuuksia. Analyysin kohteena voi olla yksi idea tai kyseessä voi olla monen idean välinen vertailu. Analyysi tarkoittaa keinotekoisessa (kontrolloidussa, simuloidussa, keksityssä...) ympäristössä suoritettavaa tarkastelua erotuksena käytännön tilanteen tarkastelusta (jota kuvailee raportti). Analyysi on käytännössä esimerkiksi idean ominaisuuksien pohdinta, kokeen tulosten raportointi tai eri ideavaihtoehtojen vertailutesti.

*Raportti.* Raportti kuvailee tai analysoi käytännön tilanteen. Toisin kuin analyysi, raportti sisältää tietoa käytännön toteutuksesta tai toteutuksista todellisissa käyttötilanteissa. Käytännössä raportti on esimerkiksi jotain tilannetta, ilmiötä tai prosessia tarkkailtaessa tehtyjen huomioiden kuvaus, esimerkiksi haastattelun kuvaus.

*Kooste.* Kooste esittelee tietyn aihealueen ideoita. Kooste itsessään ei sisällä uusia ideoita eikä uutta tietoa esitettävistä ideoista. Sen sijaan kooste voi sisältää lisämateriaalia esimerkiksi ideoiden ymmärtämisen helpottamiseksi. Kooste on käytännössä esimerkiksi oppimateriaali, katsaus tai esittely. Yksinkertaisimmillaan kooste pelkästään kokoaa ideoita yhdeksi loogiseksi kokonaisuudeksi, jolloin voidaan puhua 'linkkilistasta'.

*Luokittelematon.* Tähän luokkaan kuuluvat ideat, jotka eivät kuulu yhteenkään edellä mainituista luokista. Idea voi perustellusti olla luokittelematon ainakin kolmesta syystä. Yksi syy on se, idean tarkoitus ei kunnolla vastaa mitään vaihtoehtoa. Toinen

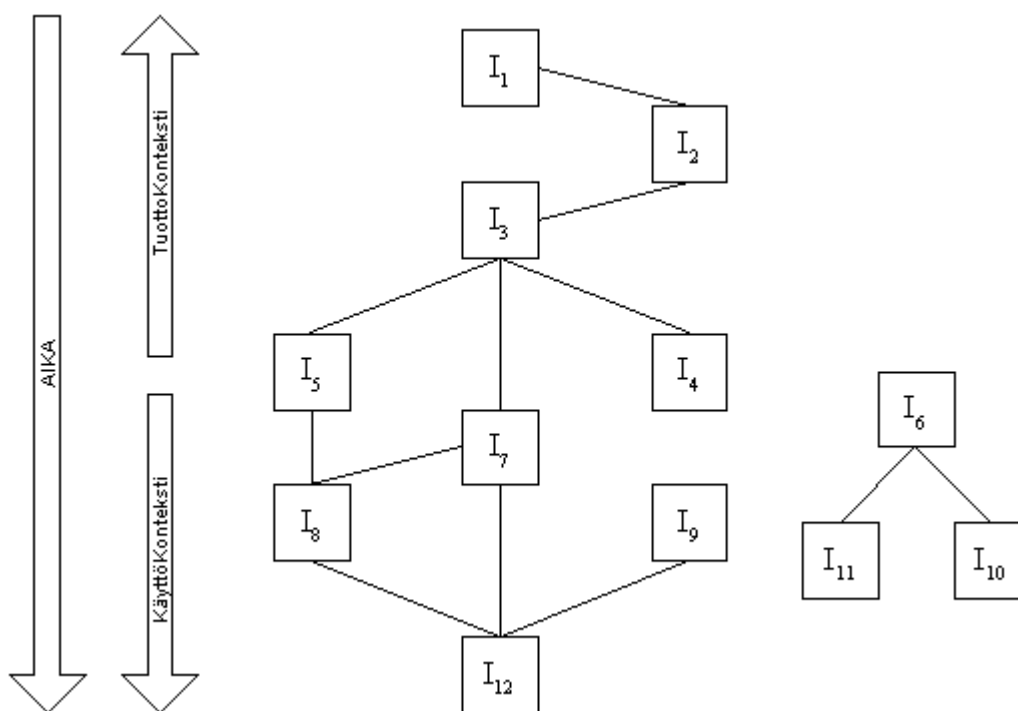
syy on se, että luokkaa ei tiedetä. Järjestelmässä saatetaan esimerkiksi julkaista ideoita, joiden tuottajilta ei voida kysyä idean tarkoitusta. Kolmas syy on se, että idean tarkoitusta ei haluta ilmoittaa. Idean tarkoitus saatetaan jättää ilmoittamatta esimerkiksi silloin, kun ideaa ei haluta arvioitavan minkään erityisen luokan perusteella. Tällaisia voivat olla esimerkiksi yksittäiset mielipiteet ja neuvot.

Luokittelemattomien ideoiden luokka ei ole olemassa sitä varten, että sitä kautta voitaisiin julkaista kaikentyyppistä informaatiota, vaan luokittelemattomankin tiedon tulisi olla 'ideamaista'. Ideat ovat toki vain tiedon joukon pieni osajoukko. Muita merkittäviä tiedon lajeja ovat ainakin erilaiset päätökset, kuten säännöt ja lait, sekä havaintoaineisto eli data ja tilastot. Epäilemättä sellainen idean kontekstin kuvaus olisi hyödyllinen, joka sisältäisi viittauksia myös tällaiseen tietoon. Universaalia tiedon välittymisen mallia ja järjestelmää, jossa käsiteltäisiin kaikkia tiedon lajeja, ei tämän tutkielman puitteissa ole mahdollista kehittää. Siksi tässä ja jatkossa oletetaan, että järjestelmässä käsitellään vain ideoita, eikä muuntyyppistä tietoa ole käytettävissä.

## 4.2 Ideoiden väliset suhteet

Ideoiden välisten suhteiden kuvaaminen on tärkein ja tehokkain keino kuvata idean kontekstia. Ideoiden välillä voi vallita suoraan linkkien avulla ilmaistuja suhteita ja epäsuoria yhteyksiä esimerkiksi sen perusteella, että ne on julkaistu samoihin aikoihin tai että niillä on yhteisiä tuottajia, esityksiä tai julkaisijoita. Tässä käsitellään suoraan linkkien avulla muodostettavia suhteita.

Idean kontekstiin kuuluu ideoiden välisten linkkien perusteella kahdentyyppisiä ideoita. Tuottokontekstiin kuuluvat ideat, joita on käytetty idean lähteenä eli ne ideat, joihin viitataan. Käyttökontekstiin kuuluvat ne ideat, jotka käyttävät ideaa lähteenä eli ideaan viittaavat ideat. Kuvassa 3 on esimerkkitilanne ideoiden välillä vallitsevista suhteista.



Kuva 3. Ideoiden väliset suhteet

Tuottokontekstiin kuuluvat ideat ovat aina ideaa itseään aikaisemmin tuotettuja ja käyttökontekstiin kuuluvat ideat myöhemmin tuotettuja. Järjestelmän on huolehdittava siitä, että tuottokontekstiin voidaan liittää vain ideoita, joiden tuottoajankohta on tuotettua ideaa aiempi. Välittömään tuottokontekstiin kuuluvat ne ideat, jotka on linkitetty ideaan ja välilliseen ne, jotka kuuluvat jonkin välittömään tuottokontekstiin kuuluvan idean välittömään tai välilliseen tuottokontekstiin. Vastaavasti erotetaan välittömään ja välilliseen käyttökontekstiin kuuluvat ideat. Ideat ovat ajan suhteen totaalisesti järjestettyjä ja välilliseen tuottokontekstiin (käyttökontekstiin) kuuluvat ideat on tuotettu aidosti aikaisemmin (myöhemmin). Kuvassa 3 ylempänä sijaitsevat ideat on tuotettu aikaisemmin kuin alempana sijaitsevat. Selvyiden vuoksi ideat on kuvassa indeksoitu niiden tuottoajankohdan perusteella.

Informaatioavaruuteen muodostuu linkkien ansiosta verkko, jonka solmuja ovat ideaa kuvaavat metatietotietueet ja kaaria ideoita yhdistävät linkit. Linkit ovat aina kaksisuuntaisia eli niitä voidaan edetä kummastakin linkin yhdistämästä solmusta käsin. Toisaalta kaarista tiedetään aina, ovatko ne tuottokontekstin linkkejä, jotka

osoittavat edeltäjiin vai käyttökontekstin linkkejä, jotka osoittavat seuraajiin. Siten verkon voi hahmottaa koostuvan kahdesta suunnatusta verkosta. Nämä suunnatut verkot ovat ideoiden tuotto- ja käyttöjärjestyksen ansiosta syklittömiä, mikä on automaattisen käsittelyn kannalta erittäin hyödyllinen ominaisuus.

Verkon kaikki solmut eivät käytännössä ole kytköksissä toisiinsa, vaan verkko on osaverkkojen joukko. Jokaisessa osaverkossa on vähintään yksi paikallinen minimi eli idea, jolla ei ole tuottokontekstia ja yksi paikallinen maksimi eli idea, jolla ei ole käyttökontekstia. Esim. kuvassa 3 on kaksi osaverkkoa ja paikallisia minimejä ovat  $I_1$ ,  $I_6$  ja  $I_9$  ja paikallisia maksimeja  $I_4$ ,  $I_{10}$ ,  $I_{11}$  ja  $I_{12}$ . Käytännössä ensimmäiset järjestelmässä julkaistut ideat (ainakin ensimmäinen) ovat paikallisia minimejä ja uusimmat (ainakin uusin) ideat paikallisia maksimeja.

Perinteisesti idean kontekstista on helposti nähtävissä lähdeluettelon avulla vain välittömään tuottokontekstiin kuuluvat ideat. Esimerkiksi  $I_3$ :n kontekstista olisi tällöin käytettävissä vain  $I_2$ . Digitaalisessa järjestelmässä linkkien kaksisuuntaisuuden ansiosta myös välittömään käyttökontekstiin kuuluvat ideat tulevat välittömästi käytettäviksi. Esimerkiksi  $I_3$ :n välitön käyttökonteksti muodostuu ideoista  $I_4$ ,  $I_5$  ja  $I_7$ . Luonnollisesti myös välillistä tuotto – ja käyttökontekstia päästään linkkien ansiosta tehokkaasti käsittelemään.

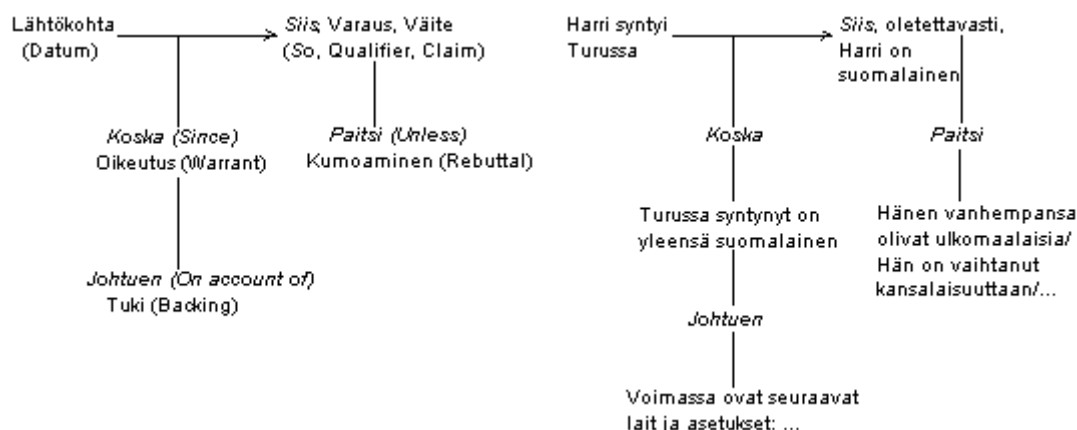
Linkkien kuvaaminen luo jo sinällään mahdollisuuksia ideoiden löytämisen ja valinnan tukemiseksi. Idean arviointi voi kuitenkin tällöin perustua vain linkkien lukumäärien ja niiden luomien assosiaatioiden tarkasteluun. Arviointimahdollisuuksien laajentamiseksi tulisi myös linkin merkityksestä tietää jotakin. Lähteen merkitys on usein pääteltävissä siitä, miten ja missä yhteydessä lähettä idean esityksessä käytetään. Merkityksen automaattinen päättely tämän tiedon perusteella on hyvin hankalaa ja se vaatisi esityksen sisällön analysoinnin menetelmiä. Vaihtoehtona on ratkaisu, jossa linkin muodostaja eli idean tuottaja eksplisiittisesti määrittelee linkin merkityksen.

Lähteiden käyttöä ja viittausten merkitystä todellisissa tilanteissa on jonkin verran tutkittu [CaH00, Gar96, WhW97]. Näiden tutkimusten valossa näyttää siltä, että erilaisia syitä viittaamiseen on huomattavan paljon, käytännöt vaihtelevat yksilöittäin

ja aihealueittain ja lähteiden käyttäjilläkään ei ole usein selkeää käsitystä ratkaisujensa syistä. Koska lähteiden joukossa on useimmiten paljon lähteitä, jotka eivät ole aivan olennaisia ja jokaisen lähteen metatietomuotoinen kuvaaminen lisää idean tuottajan työmäärää, kannattaa itse esitykseen kuuluvat ja metatietokuvaukseen kuuluvat viittaukset erottaa toisistaan. Tällöin metatietokuvauksessa riittää esitellä vain merkityksellisimmät viittaukset. Ratkaisu mahdollistaa idean linkittämisen myös sellaisiin ideoihin, joihin esityksessä ei lainkaan viitata. Näin linkkien käytön ei tarvitse millään tavalla vaikuttaa siihen tyyliin, jolla idea esitetään ja siihen, miten viittauksia itse esityksessä käytetään. Ratkaisu varmistaa myös sen, että dokumenttien käytettävyys ei riipu siitä, onko metatiedonvälitysjärjestelmä käytettävissä vai ei.

Tiedon tuottoprosessin mallintamiseen soveltuvia järjestelmiä kehitettäessä on luotu ja käytetty erilaisia malleja yksittäisten tietoalkioiden (väitteiden ja argumenttien) välisten suhteiden kuvaamiseksi. Näissä malleissa linkkityypit on usein määritelty teoreettisesta näkökulmasta, mikä luonnollisesti johtaa hieman erilaisiin tuloksiin kuin todellista lähteiden käyttöä koskevat analyysit. Argumentoinnin kuvaaminen poikkeaa ideoiden kuvaamisesta toki siinä mielessä, että yksittäinen argumentti on jollain tavalla pienempi kokonaisuus kuin yksittäinen idea. Tiedon kehittymisen prosessi lienee mittakaavaeroa lukuunottamatta kuitenkin hyvin samantapainen. Siksi näissä järjestelmissä käytettyjen linkkityyppien tarkastelu on hyödyllistä. Tässä käsitellään kahta tällaista mallia, jotka ovat Toulminin kehittämä Toulminin argumenttimalli [NeM92, Low85] ja Rittelin kehittämä IBIS –metodi [CoB88] (IBIS sanoista Issue Based Information System).

Kuvassa 4 on esitetty Toulminin argumenttimalli siten, että vasemman puoleinen kuva esittää yleistä rakennetta ja oikeanpuoleinen esimerkkiä. Argumenttien väliset suhteet on esitetty nimeämällä argumentit (Lähtökohta, Varaus, Väite, Oikeutus, Tuki ja Kumoaminen), ei tyypittämällä itse linkkejä. Linkkien perustulkinta on kuvissa kuitenkin esitetty kursivoidulla tekstillä (*Siis, Koska, Johtuen ja Paitsi*).



Kuva 4. Toulminin argumenttimalli [NeM92]

Toulminin argumenttimallin tarkoituksena on mallintaa väitteeseen johtanutta päättelyprosessia. Siksi malli on luotu väitteen näkökulmasta, eikä malli ota kantaa siihen, missä järjestyksessä argumentit linkitetään toisiinsa. Täten mallissa voidaan tunnistaa neljä linkkityyppiä. Lähtökohdan ('Harri syntyi Turussa') ja väitteen ('Harri on suomalainen') yhdistävä linkki (*Siis*) tarkoittaa sitä, että väite **olettaa** lähtökohdan olevan tosi. Oikeutuksen ('Turussa syntynyt on yleensä suomalainen') ja väitteen yhdistävä linkki (*Koska*) tarkoittaa sitä, että väite **perustuu** oikeutukseen. Oikeutuksen tuen ('Voimassa ovat seuraavat lait ja asetukset: ...') ja oikeutuksen yhdistävä linkki (*Johtuen*) tarkoittaa sitä, että oikeutuksen tuki **tukee** oikeutusta. Varauksen ('oletettavasti') ja kumoamisen ('Hänen vanhempansa olivat ulkomaalaisia') yhdistävä linkki (*Paitsi*) tarkoittaa sitä, että kumoaminen **horjuttaa** väitettä eli vähentää sen yleistä pätevyyttä.

IBIS-metodissa mahdollisia linkkityyppejä on kahdeksan: **yleistää** (generalizes), **erikoistaa** (specializes), **korvaa** (replaces), **kyseenalaistaa** (questions), **tukeutuu** (is-suggested-by), **kannattaa** (supports), **vastustaa** (objects-to) ja **ottaa kantaa** (responds-to). Kuten Toulminin mallissakin, linkkien merkitys on määritelty vain toiseen suuntaan eli viittaavasta viitattuun.

Esitettyjen kahden linkkityypistön linkit voidaan jakaa kahteen ryhmään. Ensimmäiseen ryhmään kuuluvat linkit, joissa viittaava idea jollain tavalla käyttää hyväkseen viitattua ideaa. Nämä linkit ovat olettaa, perustuu, yleistää, erikoistaa ja

tukeutuu. Toiseen ryhmään kuuluvat ne, joissa viittaava idea jollain tavalla kohdistuu viitattuun ideaan. Nämä ovat tukee, horjuttaa, korvaa, kyseenalaistaa, kannattaa, vastustaa ja ottaa kantaa. Kutsutaan ensimmäisen ryhmän linkejä lähdelinkeiksi ja jälkimmäisiä kohdelinkeiksi.

Lähdelinkit voidaan jaotella sen mukaan, kuinka tärkeä lähde on, eli kuinka vahvasti viittaavan idean pätevyys riippuu viitattusta ideasta. Yksinkertaisimmillaan voidaan ryhmä jakaa kahteen joukkoon, eli vahvaan riippuvuuteen (olettaa, perustuu) ja löyhempään suhteeseen (yleistää, erikoistaa, tukeutuu). Vahvan riippuvuuden kyseessä ollessa viittaava idea perustuu viitattavaan ideaan, jolloin viitattavan idean arvosta voidaan päätellä viittaavan idean arvoa ja päin vastoin. Löyhemmän riippuvuuden kohdalla tällaista vahvaa korrelaatiota ei voi olettaa. Jotta idean tuottajalla olisi mahdollisuus ilmaista viittauksen luonne ja jotta se voidaan päättelymenetelmiä käytettäessä ottaa huomioon, on lähdelinkit syytä jakaa kahteen osajoukkoon. Täten saadaan seuraavat tavat viitata lähteenä käytettyihin ideoihin:

**Perustavat ideat:** ideat, jotka perustavat viittaavaa ideaa

**Taustoittavat ideat:** ideat, jotka taustoittavat viittaavaa ideaa

Kohdelinkit puolestaan jakautuvat luontevasti kolmeen osajoukkoon, jotka ovat viitattua tukevat (tukee, kannattaa), viitattua horjuttavat (horjuttaa, korvaa, kyseenalaistaa, vastustaa) ja neutraalit (ottaa kantaa). Tästä saadaan kolme tapaa viitata kohteena käytettyihin ideoihin:

**Tarkasteltavat ideat:** ideat, joita viittaavassa ideassa tarkastellaan

**Tuetut ideat:** ideat, joita viittaava idea tukee

**Horjutetut ideat:** ideat, joita viittaava idea horjuttaa

Edellä linkkien merkitys on määritelty viittaavan idean eli tuottokontekstin näkökulmasta. Jotta viittauksia voitaisiin käyttää käyttökontekstin luomiseksi, on linkkien merkitys määriteltävä myös viitattavan idean näkökulmasta. Nimetään linkit viitattavan idean näkökulmasta seuraavasti (vastaavassa järjestyksessä kuin edellä):

**Perustuvat ideat:** ideat, jotka perustuvat viitattavaan ideaan

**Taustoittuvat ideat:** ideat, joita viitattava idea taustoittaa

**Tarkastelevat ideat:** ideat, jotka tarkastelevat viitattavaa ideaa

**Tukevat ideat:** ideat, jotka tukevat viitattavaa ideaa

**Horjuttavat ideat:** ideat, jotka horjuttavat viitattavaa ideaa

Kaiken kaikkiaan linkkityyppjä on siis viisi, mutta koska tuottokontekstiin kuuluvilla ideoilla on selvästi eri merkitys kun käyttökontekstiin kuuluvilla ideoilla, voi yhteen ideaan suoraan liittyä muita ideoita kymmenessä eri merkityksessä.

Valmiita sääntöjä linkkien käytölle on vaikea asettaa. Perusrajoitus on, että yhden idean voi linkittää toiseen vain yhdellä linkkityypillä. Tämä rajoitus on syytä tehdä yksinkertaisuuden nimissä ja linkkien määrän rajoittamiseksi. Seuraavat huomiot on tulkittava enemmän ideatyypien ja linkkityypien selventäjinä kuin tiukkoina sääntöinä.

Useimpiin ideoihin liittyy idea tai joitakin ideoita, johon se perustuu. Miksi-ideat perustunevat todennäköisimmin raportteihin, mitä-ideat miksi-ideoihin, miten-ideat mitä-ideoihin, ja tulokset, analyysit, raportit ja koosteet tuloksiin ja miten-ideoihin. Analyysien kohdalla perusteena käytettyjen ideoiden esittely on erityisen tärkeää. Koosteen perustuminen johonkin ideaan tarkoittaa sitä, että koosteen tyyli perustuu johonkin (esim. pedagogisiin näkökulmiin), koska itse kooste ei sisällä uutta ideaa. Samoin raportin perustuminen johonkin on tulkittava niin, että raportin tuottamisessa käytetyt menetelmät perustuvat johonkin ideaan, koska itse raportin sisältö perustuu johonkin todellisuudessa vallinneeseen tilanteeseen, jonka kuvaaminen juuri on raportin tarkoitus.

Ideaan tulee liittää sitä taustoittavia ideoita silloin, kun idea on esitetty erityisen kontekstisidonnaisessa muodossa eli kun idean oikea ymmärtäminen vaatii tiettyä suhtautumistapaa tai näkökulmaa. Toinen tärkeä syy taustoittuu -linkin käyttöön on se, että tämän avulla voidaan viitata vaihtoehtoihin näkökulmiin. Koska ideat ovat enemmän tai vähemmän kontekstisidonnaisia, useimpiin ideoihin liittyy taustoittavia ideoita. Erityisesti koosteet ovat tärkeitä taustoittajia.



Tarkastelee-linkkiä tulee käyttää silloin, kun idean päätarkoituksena on tarkastella jotakin toista ideaa. Ideaan ei siis tule tämän linkkityypin avulla linkittää kaikkia ideoita, joita esityksessä tarkastellaan, vaan pelkästään ne ideat, joiden tarkasteleminen on idean päätehtävänä. Mitä-, ja miten-, ja tulos -tyyppisissä ideoissa tarkastelee -linkin käyttö lienee siten harvinaista, miksi- ja analyysi-tyyppisissä yleisempää ja raportti- ja kooste -tyyppisissä ideoissa kaikkein yleisintä. Analyysien, raporttien ja koosteiden tulisi aina kohdistua yhteen tai useampaan ideaan, joko tarkastelee-, tukee-, tai horjuttaa -linkkien kautta. Koostetyyppisissä ideoissa tarkastelee -linkkien olemassaolon tulisi olla itsestään selvää, koska kooste kohdistuu koostettaviin ideoihin, eikä koostettavia ideoita pidä koosteessa tukea eikä horjuttaa.

On selvää, että käytännössä kaikissa ideoiden esityksissä otetaan jollain tavalla kantaa toisten ideoiden arvoon. Jopa koosteissa voi esiintyä muita ideoita koskevia mielipiteitä ja arvotuksia. Tukee ja horjuttaa -linkkejä tulisi kuitenkin käyttää silloin ja vain silloin, kun idean ensisijainen tarkoitus on tarkastella toista ideaa tai ideoita ja tarkastelu tuottaa selvän loppupäätelmän. Ainoastaan analyysityyppisten ideoiden kohdalla näiden linkkien käyttö on itsestään selvää, koska analyysin tulisi paitsi tarkastella ideoita, myös tuottaa niitä koskevaa uutta tietoa. Myös miksi -tyyppiset ideat saattavat usein tuottaa perusteltuja näkökulmia, jotka tukevat tai horjuttavat muita ideoita.

Lopuksi on korostettava sitä, että linkkityypit on tarkoitettu ainoastaan ideoiden välisten suhteiden kuvaamiseen. Linkkiä ei tule käyttää, jos idean lähteenä tai kohteena on ollut esimerkiksi esitys, tuottaja, julkaisija tai ideoiden välisiä suhteita kuvaava linkki.

### 4.3 Julkaisukonteksti

Idea on esitettävä sellaisessa muodossa, jossa se voidaan siirtää, tallentaa ja ymmärtää. Jotta esitys olisi käytettävissä, se on tallennettava ja asetettava saataville johonkin tietovarastoon. Idean julkaisukonteksti kuvaa idean esitystä ja esitysten julkaisuja.

Esityskohtaisen metatiedon luominen on esitysten tuottajien tehtävä ja julkaisukohtaisten tietojen luominen on julkaisijoiden tehtävä.

Jokainen julkaisutapahtuma koskee vähintään yhtä esitystä, mutta julkaistava dokumentti voi olla myös monen esityksen kokoelma, kuten aikakauslehden yksi numero, joka sisältää monta artikkelia. Jotta dokumentti voitaisiin koostaa monesta eri esityksestä, tulee jokaisella esityksellä olla yksikäsitteinen tunniste. Tämän attribuutin nimi olkoon **esityksen tunniste**. Vastaavasti kuin ideoistakin, esityksistäkin on hyvä tietää niiden nimi, tuottajat ja tuottoajankohta. Näiden attribuuttien nimet olkoot **esityksen nimi**, **esityksen tuottajat** ja **esityksen tuottoajankohta**.

Esitykseen on tietenkin kuuluttava viittaukset niihin ideoihin, jotka esityksessä esitetään. Tämän attribuutin nimi olkoon **esityksen ideat**. Linkin avulla jokaiseen ideaan muodostuu lista niistä esityksistä, joissa idea esitetään. Tämän ideakohtaisen attribuutin nimi olkoon **idean esitykset**. Tämä on siis se attribuutti, joka kytkee idean sen julkaisukontekstiin.

Julkaisijat saattavat haluta tarjota tiedon tuottajille ja tarvitsijoille muitakin palveluita kuin dokumenttien ja metatiedon hallintaa. Esimerkkejä tällaisista palveluista ovat erilaiset viestintäkanavat, joita käsiteltiin luvussa 3.2, julkaisijan esittely (esim. linkki 'kotisivulle') ja julkaisijan kontaktitiedot, kuten sähköpostiosoite. Tiedot tällaisista palveluista voidaan koota yhden julkaisijakohtaisen attribuutin alle. Tämän attribuutin nimi olkoon **viestintäkanavat**. Jotta viestintäkanavien käyttö olisi helppoa, ne tulisi jollakin tavalla luokitella ja niiden tyyppi esitellä. Koska viestintäkanavat ovat kuitenkin ideoidenvälitysjärjestelmän ulkopuolisia erillisiä palveluita, kanavien luokittelua ja toteuttamista ei tässä tarkemmin käsitellä.

Edellä mainittujen attribuuttien lisäksi esityksiin voisi liittyä hyvinkin monipuolisesti esitystä ja sen sisältöä kuvaavaa tietoa. Esimerksi Dublin Coren elementeistä esityksen kuvaamiseen sopisivat edellä mainittuja attribuutteja vastaavien lisäksi erityisen hyvin aihe, kuvaus, laji ja kieli.

Kun ideasta on tuotettu esitys, se voidaan julkaista. Jokaista julkaisutapahtumaa vastaa yksi dokumenttiavaruuden dokumentti. Jokaista dokumenttia vastaa

metatietoavaruudessa täsmälleen yksi metatietotietue. Siten metatietotietueen tunniste toimii aina yhden julkaisutapahtuman tunnisteena. Tämän attribuutin nimi olkoon **julkaisun tunniste**. Jotta metatietotietue voidaan yhdistää sitä vastaavaan dokumenttiin, on kuvaukseen kuuluttava dokumentin tunniste. Olkoon tämän attribuutin nimi **dokumentin tunnisteet** (dokumentilla voi olla vaihtoehtoisia tunnisteita eri dokumenttienvälitysjärjestelmiä varten). Julkaisu on myös linkitettävä niihin esityksiin, joita julkaisu koskee. Tämän attribuutin nimi olkoon **dokumentin esitykset**. Tämän linkin avulla jokaiseen esitykseen muodostuu lista niistä julkaisutapahtumista, joita esitykseen liittyy. Tämän esityskohtaisen attribuutin nimi olkoon **esityksen julkaisut**.

Jokainen dokumentti on tallennettuna dokumenttiavaruudessa johonkin dokumenttivarastoon ja jokainen metatietotietue on tallennettu dokumenttivarastoa vastaavaan metatietovarastoon. Siten metatietovarasto, jossa tietty metatietotietue on tallennettuna, kertoo suoraan dokumentin julkaisijan. Julkaisutapahtumien ajankohdat on syytä tallentaa, jotta ideoiden saatavillaoloajan pituus voidaan ottaa huomioon ideoita arvioitaessa. Tämän attribuutin nimi olkoon **dokumentin julkaisuaikajankohda**. Muista hyödyllisistä dokumenttia ja julkaisua kuvaavista attribuuteista mainittakoon Dublin Coren elementti formaatti, jonka avulla voidaan ilmaista dokumentin fyysinen tai digitaalinen ilmiasu.

Tiettyyn tietovarastoon kuuluvat dokumentit ovat tietyn julkaisijan julkaisemia. Jokaisesta ideasta tiedetään, ketkä idean ovat julkaisseet ja jokaisesta idean julkaisijasta tiedetään, mitä muita ideoita he ovat julkaisseet. Julkaisijat eli metatietokannat on luokiteltava, jotta näitä tietoja päästään tehokkaasti hyödyntämään.

Henkilökohtaisia tietovarastoja tarvitaan jo sen takaamiseksi, että kaikilla järjestelmän käyttäjillä olisi yhtäläiset mahdollisuudet julkaista materiaalia. Henkilökohtaisessa tietovarastossa julkaistaan pelkästään dokumentteja, joiden sisältämien ideoiden tai esityksien tuottamiseen henkilö on osallistunut. Siten henkilökohtaisen tietovaraston kautta saadaan helposti selville, mitä muita ideoita tai esityksiä jonkin idean tai esityksen tuottaja on tuottanut. Jotta ideasta päästään käsiksi idean tuottajien henkilökohtaisiin tietovarastoihin, on tietovarastot lueteltava tunnisteidensa avulla. Tämän ideakohtaisen attribuutin nimi olkoon **idean tuottajatiedot**. Vastaavasti

esityksistä tarvitaan linkit esitysten tuottajien tietovarastoihin. Tämän esityskohtaisen attribuutin nimi olkoon **esityksen tuottajatiedot**.

Henkilökohtaisten tietovarastojen lisäksi tarvitaan tietovarastoja, jotka ovat julkaisijakohtaisia, ts. tietovarastoja, joiden ylläpitäjä ei ole itse tuottanut ideaa tai esitystä. Mitä useampaan ryhmään julkaisijat luokitellaan, sitä monipuolisemmin julkaisijatietoja voidaan hyödyntää. Seuraavassa esitellään kolme julkaisijatyyppeä, jotka ovat arkistoivat, organisaatiokohtaiset ja yhteisökohtaiset tietovarastot.

*Arkistoivat tietovarastot.* Henkilökohtaisessa tietovarastossa voidaan määritelmän mukaan julkaista vain ideoita ja esityksiä, joiden tuottamiseen henkilö on osallistunut. Pelkästään henkilökohtaisiin tietovarastoihin perustuvassa järjestelmässä ei voida julkaista dokumentteja, joiden tuottajiin ei liity henkilökohtaisia tietovarastoja. Henkilökohtaisten tietovarastojen toiminta ja ylläpidon jatkuvuus ei välttämättä ole kovin varmaa, joten on vaara, että pelkästään henkilökohtaisiin tietovarastoihin tallennetut metatiedot ja dokumentit saatetaan 'kadottaa'. Tiedon tuottajia ei välttämättä motivoi julkaista tietoa sellaisessa järjestelmässä, jossa tiedon saatavuudesta ja säilyvyydestä vastaaminen on pelkästään tuottajien itsensä vastuulla. Esimerkiksi ei-digitaalisessa muodossa olevien dokumenttien kopiointi ja toimittaminen tiedon tarvitsijoille saattaa olla tiedon tuottajille ylivoimaista, mikäli ideat osoittautuvat liian kysytyiksi.

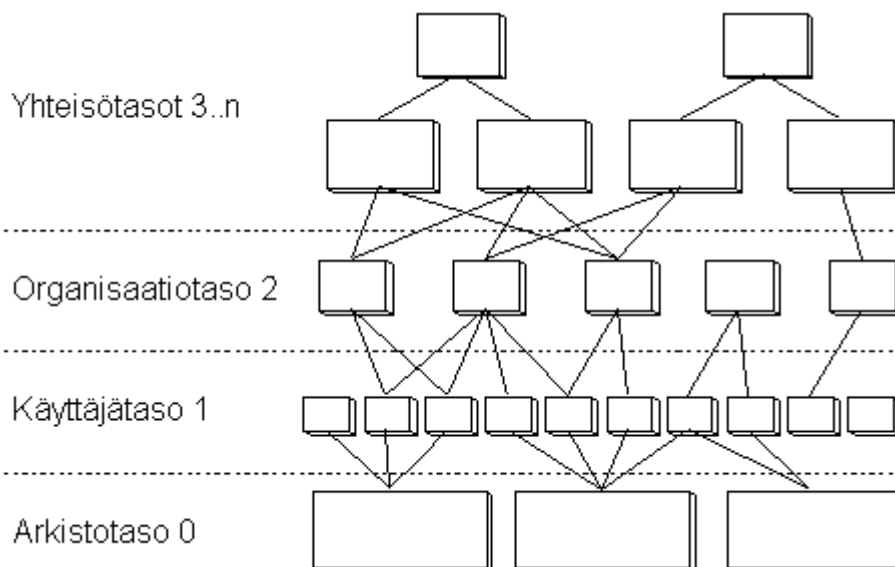
Näiden ongelmien ratkaisemiseksi tarvitaan arkistoivia tietovarastoja. Arkistoivan tietovaraston ylläpitäjän tehtävänä on julkaista sellaisia dokumentteja, joiden tuottajilla ei ole henkilökohtaisia tietovarastoja, varmistaa julkaisemansa metatiedon ja dokumenttien säilyvyys ja saatavuus ja vastata dokumenttien toimittamisesta ideoiden tarvitsijoille. Yhdessä tietovarastossa tuskin voidaan hallita kaikkia mahdollisia dokumentteja. Arkistoivat tietovarastot voivatkin keskittyä esimerkiksi tietynkielisten, tietyllä alueella tuotettujen tai tietyn tyyppisten, esim. painetussa muodossa tallennettujen dokumenttien arkistointiin. Arkistoivan tietovaraston tehtävänä ei kuitenkaan ole arvioida tai valikoida ideoita tarvitsijoiden puolesta. Siten idean julkaisemista tai julkaisemattomuutta tietyssä arkistoivassa tietokannassa ei pidä käyttää idean arvoa mittaavana kriteerinä.

*Organisaatiokohtaiset tietovarastot.* Idean julkaiseminen organisaatiokohtaisessa metatietokannassa lisää samalla tavalla idean saatavuutta ja säilyvyyttä kuin arkistoitavan metatietokannan käyttö. Nimensä mukaisesti organisaatiokohtaisessa metatietokannassa julkaistaan kuitenkin vain organisaatiokohtaisia ideoita. Julkaisusta voidaan tällöin päätellä, että organisaatio on tarkistanut ja hyväksynyt idean ja sitä koskevan metatiedon julkaisemisen omissa nimissään. Siten organisaatiokohtainen julkaisu lisää metatiedon luotettavuutta aivan eri tavalla kuin henkilökohtainen tai arkistokohtainen julkaisu. Organisaatiokohtainen tietovarasto toimii myös tiedon luokittelijana, koska yleensä tietyssä organisaatiossa tuotetut ideat ovat jollakin tavalla samantyyppisiä. Yksityiskohtaisemmin organisaatiokohtaisen julkaisijan ja julkaisemisen merkitystä käsittelevät Kling, Spector ja McKim [KSM02].

*Yhteisökohtaiset tietovarastot.* Tiedon välittäjät voivat arvioida, valita ja luokitella tietoa tiedon tarvitsijoiden puolesta. Idean arviointi ei ole arkistojen ylläpitäjien tehtävä. Se ei voi olla myöskään tiedon tuottajien tai tietoa tuottavien organisaatioiden itsensä tehtävä. Siksi tarvitaan erillinen tietovarastotyyppi, joka kuvaa tietoa arvioivia tiedon välittäjiä. Tiedon arviointi tapahtuu aina jostakin näkökulmasta käsin. Tietoa arvioiva välittäjä palveleekin jotain virtuaalista tai käytännössä toimivaa yhteisöä, joka jakaa välittäjän mielipiteet. Yhteisökohtaisessa tietovarastossa julkaistaan dokumentteja, jotka tiedon välittäjän mielestä todennäköisesti parhaiten vastaavat kohdeyhteisön tiedon tarvetta. Koska yhteisökohtaiseen tietokantaan voidaan koota ideoita monilta tuottajilta ja monista organisaatioista, yhteisökohtainen tietovarasto toimii tehokkaana tiedon kokoajana ja luokittelijana ja organisaatorajat ylittävän tiedon välittymisen tehostajana.

Avoimessa, hajautetussa ja globaalissa tiedonvälitysjärjestelmässä voi olla miljoonia tietovarastoja. Tiedon löytämisen ongelma ei tällöin ole pelkästään siinä, miten tietystä tietovarastosta löydetään hyödylliset tiedot. Yhtä tärkeää on ensin löytää ne tietovarastot, joista tietoa kannattaa etsiä. Edellä mainittujen neljän tietovarastotyypin käyttö antaa mahdollisuuden järjestää tietovarastot hierarkiaksi. Kuva 5 esittää tällaista tietokantojen hierarkiaa. Kuvassa laatikot kuvaavat metatietokantoja ja laatikoiden koko ja lukumäärä viittaavat suuntaa-antavasti metatietokantojen kokoihin ja lukumääriin suhteessa toisiinsa. Metatietokantoja yhdistävä viiva tarkoittaa sitä, että kumpikin julkaisija on julkaissut ainakin yhden samaa ideaa koskevan

dokumentin. Kuvassa on esitetty joitain mahdollisia suhteita, joita näin muodostuu metatietokantojen välille.



Kuva 5. Tietovarastojen hierarkia

Tasolle nolla kuuluvat arkistoitavat tietovarastot, tasolle yksi henkilökohtaiset tietovarastot, tasolle kaksi organisaatiokohtaiset tietovarastot ja tasoille kolmesta ylöspäin yhteisökohtaiset tietovarastot. Hierarkia muodostetaan asettamalla periaate, että ylemmällä tasolla voidaan julkaista tietoa vain, kun se on ensin julkaistu alemmilla tasoilla. Aluksi dokumentti voidaan julkaista joko tasolla nolla tai tasolla yksi. Tämän jälkeen dokumentti voidaan julkaista tasoilla nolla, yksi tai kaksi. Kun dokumentti on kerran julkaistu tasolla kaksi, se voidaan tämän jälkeen julkaista tasoilla nolla, yksi, kaksi tai kolme ja niin edelleen. Jotta myös organisaatioiden ulkopuolella tuotettuja ideoita voitaisiin julkaista yhteisötasolla, on tällaisten ideoiden tuottajien tarvittaessa perustettava uusia organisaatioita tai käytettävä sellaista organisaatiokohtaista tietovarastoa, jonka ylläpitäjä on valmis julkaisemaan kaiken sille tarjotun materiaalin. Avoimessa järjestelmässähän organisaatiokohtaisten tai muunkaan tyyppisten tietovarastojen perustamista ei mitenkään rajoiteta.

Järjestelyn seurauksena tietovarastojen välille muodostuu puumainen rakenne – ylemmillä tasoilla on vähemmän tietovarastoja kuin alemmilla tasoilla ja ylemmillä

tasoilla julkaistaan aina käytännössä vähemmän dokumentteja kuin alemmilla tasoilla. Ajatuksena on, että ylimmän tason yhteisöt muodostavat hierarkiaan hallittavissa olevan määrän juuria, joista tietovarastojen selaaminen voidaan tarvittaessa aloittaa. Jotta hierarkia olisi skaalautuva, ei yhteisötasojen lukumäärää ole rajoitettu, vaan uusi yhteisö voidaan perustaa ylemmälle tasolle aina, kun alemmilla tasoilla alkaa olla riittävästi sellaista materiaalia tarjolla, joka kannattaa uudessa yhteisökontekstissa julkaista.

Jotta tietovarastojen tyypit voidaan erottaa toisistaan, ne tulee merkitä metatietotietueisiin. Tietovaraston tyyppinä voidaan suoraan käyttää tason numeroa eli esimerkiksi organisaatiokohtaisen tietovaraston tyyppinumero on kaksi. Attribuutin nimi olkoon **julkaisijan tyyppi**.

#### 4.4 Metatietokuvauksen sisältö

Metatietokuvaus tallennetaan metatietotietueisiin. Metatietokuvauksen sisältö voidaankin määritellä määrittelemällä metatietotietueen sisältämät attribuutit. Yksi metatietokuvaus koostuu dokumentin julkaisua, dokumentin sisältämiä esityksiä ja esityksien sisältämiä ideoita kuvaavista attribuuteista ja niiden arvoista. Tässä metatietokuvauksen sisältö esitetään kootusti edellisiin lukuihin perustuen. Metatietotietueen koostumuksen esittely aloitetaan käymällä läpi mahdolliset arvotyyppit, joita attribuuteilla voi olla.

*Tunniste.* Tunniste identifioi informaatioavaruuden resurssin siten, että resurssiin voidaan viitata toisista resursseista käsin. Viittauksen avulla voidaan informaatioavaruudesta löytää ja hakea resursseja koskevaa tietoa. Jokainen resurssi kuuluu johonkin resurssiluokkaan. Luokilla on omat tunnistetyypinsä. Resurssiluokkia ovat ideat, esitykset, julkaisut, metatietokannat, palautekanavat ja dokumentit. Näiden tunnisteista käytetään seuraavia lyhennysmerkintöjä: idean tunniste = itunniste, esityksen tunniste = etunniste, julkaisun tunniste = jtunniste, metatietokannan tunniste = mtktunniste, viestintäkanavan tunniste = vktunniste ja

dokumentin tunniste = dtunniste. Esimerkki attribuutista, jonka arvon tyyppi on tunniste, on idean tunniste.

*Merkkijono.* Merkkijono kuvaa mitä tahansa merkeistä muodostettua arvoa. Näiden arvojen sisällöllä ei ole merkitystä järjestelmän tietojenkäsittelyllisen toiminnallisuuden kannalta. Siten merkkijonojen sisältö on aina tiedon syöttäjän vapaasti valittavissa. Idean nimi on esimerkki merkkijonotyyppisestä attribuutista.

*Merkkijonojoukko.* Merkkijonojoukko on monen merkkijonon kokoelma. Merkkijonojoukko –arvoisia attribuutteja on mallissa kaksi, idean tuottajat ja esityksen tuottajat.

*Aika.* Aika ilmoittaa päivämäärän ja se voi ilmoittaa lisäksi kellonajan. Idean tuottoajankohdan kohdalla kellonajan ilmoittaminen on pakollista silloin, kun halutaan viitata ideaan, jonka tuottoajankohta olisi muutoin sama. Näin varmistetaan se, että ideoista voidaan viitata vain aikaisemmin tuotettuihin ideoihin. Idean tuottoajankohdan lisäksi muut aika-arvoiset attribuutit ovat esityksen tuottoajankohta ja dokumentin julkaisuajankohta.

*Tyyppi.* Kun attribuutti on tyyppi-arvoinen, attribuutin arvo on valittava valmiista vaihtoehtojoukosta. Koska vaihtoehtoisten tyyppien joukko on aina numeroituva, tyyppi-arvoiset arvot voidaan esittää kokonaislukujen avulla. Idean tarkoitus valitaan väliltä 1-8 (kts. luku 4.1) ja julkaisijan tyyppi väliltä 0-n (kts. luku 4.3). Muita tyyppi-arvoisia attribuutteja ei malliin kuulu.

*Linkki (Tunniste).* Linkki on viittaus resurssista toiseen resurssiin. Linkin arvona on viitattavan resurssin tunniste. Tietty linkki-arvoinen attribuutti viittaa vain tiettyntyyppisiin resursseihin, joten tunnisteeseen tyyppi voidaan ilmoittaa attribuutin yhteydessä. Esimerkiksi linkki (mtktunniste) tarkoittaa sitä, että viittauksen kohde on metatietokanta. Julkaisijan tunniste on mallissa ainut linkki –arvoinen attribuutti.

*Linkkijoukko (Tunniste).* Linkkijoukko on monen samaan luokkaan viittaavan linkin kokoelma. Esimerkki linkkijoukko –arvoisesta attribuutista on perustuvat ideat.



Kuvassa 6 esitetään luokiteltuna kaikki metatietokuvaukseen kuuluvat attribuutit arvotyypeineen. Suluissa on lisäksi esimerkinomaisesti mainittu joitakin Dublin Coren elementtejä niiden aseman selkiyttämiseksi.

Idea	Esitys
idean tunniste: itunniste idean nimi: merkkijono idean tuottajat: merkkijononjoukko idean tuottajatiedot: linkkijoukko(itunniste) idean tuottoajankohta: aika idean tarkoitus: tyyppi idean esitykset: linkkijoukko(etunniste)  perustavat ideat: linkkijoukko(itunniste) taustoittavat ideat: linkkijoukko(itunniste) tarkasteltavat ideat: linkkijoukko(itunniste) tuetut ideat: linkkijoukko(itunniste) horjutetut ideat: linkkijoukko(itunniste)  perustuvat ideat: linkkijoukko(itunniste) taustoittavat ideat: linkkijoukko(itunniste) tarkastelevat ideat: linkkijoukko(itunniste) tukevat ideat: linkkijoukko(itunniste) horjuttavat ideat: linkkijoukko(itunniste)	esityksen tunniste: etunniste esityksen nimi: merkkijono esityksen tuottajat: merkkijononjoukko esityksen tuottajatiedot: linkkijoukko(itunniste) esityksen tuottoajankohta: aika esityksen ideat: linkkijoukko(itunniste) esityksen julkaisut: linkkijoukko(itunniste)  (aihe, kuvaus, laji, kieli, kate,...)
	Julkaisu
	julkaisun tunniste: jtunniste julkaisijan tunniste: linkki(mtktunniste) julkaisijan nimi: merkkijono julkaisijan tyyppi: tyyppi dokumentin julkaisuajankohta: aika dokumentin esitykset: linkkijoukko(etunniste) dokumentin tunnisteet: linkkijoukko(dtunniste) viestintäkanavat: linkkijoukko(vktunniste)  (formaatti)

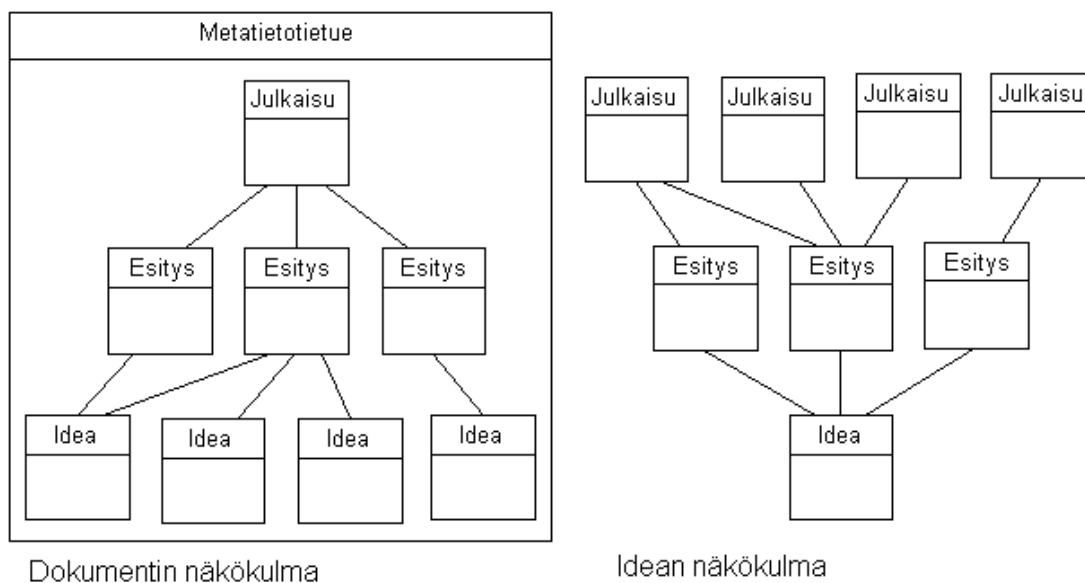
Kuva 6. Metatietotietueen attribuutit luokittain

Yksi metatietotietue sisältää kaikki dokumenttia suoraan koskevat tiedot. Jokaiseen dokumenttiin liittyy tiedot yhdestä julkaisusta, vähintään yhdestä esityksestä ja vähintään yhdestä ideasta. Esityksettömiä ideoita ei voi julkaista, kuten ei myöskään ideattomia esityksiä. Kun jokaiseen dokumenttiin liittyy täsmälleen yksi metatietotietue ja jokaiseen metatietotietueeseen täsmälleen yksi julkaisu, toimii julkaisun tunniste koko metatietotietueen tunnisteena.

Kun jo olemassaoleva idea julkaistaan uudessa esityksessä, idean tiedot kopioidaan julkaisun metatietotietueeseen. Kun julkaistaan jo olemassaoleva esitys, on kopioitava sekä esityksen tiedot että esityksen sisältämien ideoiden tiedot. Lisäksi on aina tietoa kopioitaessa päivitettävä myös kaikkia niitä metatietotietueita, jotka sisältävät

kopioitavat tiedot. Kun kyseessä on olemassaolevan tiedon julkaiseminen uudessa esityksessä, on päivitettävä idean tiedot sisältävien metatietotietueiden idean esitykset –attribuuttia lisäämällä niihin uuden esityksen tunniste ja kun julkaistaan jo olemassaoleva esitys, on päivitettävä esityksen tiedot sisältävien metatietotietueiden esityksen julkaisut –attribuuttia lisäämällä niihin uuden julkaisun tunniste. Näin menetellen jokainen metatietotietue sisältää aina tiedot kaikista dokumenttiin kuuluvista esityksistä ja esityksiin kuuluvista ideoista. Lisäksi jokaisesta metatietotietueesta päästään linkkiketjujen *dokumentin esitykset* → *esityksen ideat* → *idean esitykset* → *esityksen julkaisut* avulla käsiksi jokaiseen julkaisuun, joka sisältää samoja ideoita kuin metatietotietue. Näin metatietotietueen tietojen päivittäminen voidaan uusia dokumentteja julkaistaessa hoitaa automaattisesti.

Kuva 7 esittää metatietotietueen ja metatiedon suhdetta toisiinsa dokumenttien ja ideoiden näkökulmista käsin.



Kuva 7. Kaksi näkökulmaa kontekstitiedon muotoutumiseen.

Kuvan vasemmalla puolella on esimerkki yhden metatietotietueen sisällöstä. Yksi metatietotietue kokoaa kontekstia kuvaavat tiedot yhteen yhden dokumentin näkökulmasta käsin. Tämä näkökulma kontekstiin on sama kuin idean näkökulma

kontekstiin silloin, kun dokumentista on olemassa yksi julkaisu, julkaisu sisältää yhden esityksen ja esitys sisältää yhden idean. Tämä lienee yleisin tilanne. On myös mahdollista, että esityksellä on monta julkaisua. Koska dokumenttien ja esitysten välillä voi vallita monesta–moneen –suhteita ja myös esitysten ja ideoiden välillä voi vallita monesta–moneen –suhteita, voi kontekstin muotoutuminen olla dokumentin ja idean näkökulmista hyvin erilaistakin.

Kuvan oikealla puolella on esimerkki kontekstitiedon sisällöstä idean näkökulmasta käsin. Koska esimerkin idealla on neljä julkaisua, idean tiedot ovat kopioituna neljään metatietotietueeseen. Jokainen metatietotietue sisältää sellaista idean kontekstia koskevaa tietoa, jota mikään muu metatietotietue ei sisällä. Siten yksittäinen metatietotietue ei välttämättä määrittele yksittäisen idean koko kontekstia.

## 5 Kontekstin käyttö

### 5.1 Tiedon tarvitsijan näkökulma

Edellisissä luvuissa on käsitelty sitä, miten ideoiden luontia, julkaisua ja käyttöä kuvaavaa tietoa saadaan tallennetuksi metatietotietueisiin ja metatietokantoihin. Seuraavaksi tarkastellaan sitä, mitä hyötyä kontekstiedosta voisi tiedon tarvitsijalle olla.

Todellinen tiedonvälitysjärjestelmä koostuu lukuisista osista, kuten tietokantajärjestelmistä, tiedonsiirtoprotokollista ja laskentaa suorittavista algoritmeista. Tämän tutkielman puitteissa ei ole mahdollista tarkastella näiden toimintaa. Jatkossa oletetaan, että kontekstiedon mahdollistamat operaatiot ovat suoritettavissa, eikä operaatioiden toteutustapaan sinänsä kiinnitetä huomiota. On kuitenkin pidettävä mielessä, että hajautetussa järjestelmässä ei voida käytännössä löytää saati käsitellä kerralla kaikkia ideoita. Esimerkkinä vaikeasti löydettävissä olevista ideoista mainittakoon ideat, jotka on julkaistu vain henkilökohtaisessa tietovarastossa (ei edes missään arkistossa) ja jotka eivät viittaa ja joihin ei viitata yhdestäkään jossakin toisessa tietovarastossa julkaistusta ideasta. Ongelmana siten ei ole pelkästään ideoiden valinta, vaan myös ideoiden ja tietovarastojen löytäminen. Tarkasteluissa pääpaino on ideoiden valinnassa, mutta myös ideoiden ja tietovarastojen löytämistä sivutaan.

Tiedon löytämistä ja valintaa tukevat operaatiot perustuvat pohjimmiltaan metatietokuvausten attribuutteihin kohdistuviin kyselyihin. Mahdollisia operaatioita ovat esimerkiksi: *'hae x:n tuottamat kooste-tyyppiset ideat'*, *'hae organisaatiot, jotka ovat julkaisseet yli 50 dokumenttia'*, *'hae ne ideaa x analysoivat ideat, joista on saatavilla suomenkielinen esitys'* ja *'laske, kuinka monta henkilöä kaiken kaikkiaan on ollut tuottamassa yhteisön y julkaisemia ideoita'*. Jokainen esimerkki kuvaa erilaista kyselytyyppiä. Ensimmäisessä esimerkissä kyselyn kohteena ovat ideat, mikä ideoiden välittämiseen tarkoitettussa järjestelmässä lienee yleisin kyselyjen kohde. Toisessa esimerkissä kyselyn kohteena eivät ole ideat, vaan organisaatiot. Kontekstin mallin

perusteella tunnistettavia resurssityyppejä ovat ideoiden lisäksi myös esitykset, dokumentit, ideoiden tuottajat, esitysten tuottajat, organisaatiot, yhteisöt ja arkistot. Niinpä kyselyjä voidaan ideoiden lisäksi kohdistaa myös näihin resurssityyppeihin. Kolmannessa esimerkissä kysely on muodostettu kontekstia kuvaavan ja esityksen sisältöä kuvaavan metatiedon yhdistelmänä. Kontekstitietoa voidaan siis käyttää sisällön kuvaamiseen perustuvien menetelmien tukena. Esimerkeistä viimeisessä kyselyn kohde ei ole resurssi, vaan itse kontekstitieto. Vaikka kontekstitiedolla on ensisijaisesti vain välinearvoa eli sen tarkoituksena on helpottaa ideoiden löytämistä ja valintaa, sillä voi siis olla myös itseisarvoa. Jo nämä neljä esimerkkiä riittävät osoittamaan, että kontekstitieto on ainakin potentiaalisesti hyödyllistä.

Tiedon löytämistä ja valintaa tukevat operaatiot ovat siis pohjimmiltaan metatietokuvausten attribuutteihin kohdistuvia kyselyjä. Ongelmana on se, että tiedon tarvitsija ei useinkaan voi osata tehdä sellaisia kyselyjä, jotka tuottaisivat tiedon tarpeen kannalta parhaan tuloksen.

Kyselyjen teon vaikeuteen on kolme pääsyitä. Ensinnäkin yksinkertainenkin tiedon tarve voi johtaa attribuuttien avulla ilmaistuna hyvinkin monimutkaiseen kyselyyn. Monimutkaisuus on seurausta kontekstitiedon ilmaisuvoimasta. Joissakin tilanteissa monimutkainen kysely tuottaa tarkemman tai mielekkäämmän tuloksen kuin yksinkertaisempi versio. Toiseksi resurssien kvalitatiivisten ominaisuuksien tarkasteleminen metatietotietueiden kvantitatiivisiin ominaisuuksiin perustuen vaatii syvällistä ideoiden tuoton, julkaisemisen ja käytön prosessien ymmärtämystä, jota keskimääräisellä tiedon tarvitsijalla ei voi olla. Kolmanneksi resurssien selailu sekä niiden väleillä vallitsevien yhteyksien, suhteiden ja trendien erottaminen, hahmottaminen ja ymmärtäminen on hyvin vaikeaa pelkästään kyselyjä tekemällä ja niiden tuloksina saatavia arvoja tarkastelemalla.

Nämä ongelmat voidaan ratkaista vain kehittämällä käyttöliittymiä, jotka antavat tiedon tarvitsijalle helppokäyttöiset työkalut kontekstitiedon hyödyntämiseen. Koska kontekstitietoa tehokkaasti hyödyntävät käyttöliittymät eivät ole yleisiä tai hyvin tunnettuja, on yritettävä osoittaa, että tehokkaita käyttöliittymiä voidaan toteuttaa.

Seuraavissa luvuissa käsitellään esimerkinomaisesti kolmea kontekstitiedon käyttötapaa. Ensin tarkastellaan etenkin kyselyn tuloksen esittämistapoja, sitten tapoja kuvata tiedon tarvetta kvalitatiivisin termein ja lopuksi mahdollisuuksia suorittaa tiedon etsintää ja arviointia automaattisesti. Kaikille käyttötavoille on yhteistä se, että ne tukeutuvat vahvasti tietojenkäsittelyllisten menetelmien hyödyntämiseen. Kontekstin käytön tarkastelut voidaankin tiedon tarvitsijan näkökulmasta kiteyttää seuraavasti: kontekstia kannattaa kuvata, mikäli se yhdessä tietojenkäsittelyllisten menetelmien kanssa mahdollistaa sellaisten tehokkaiden tiedon löytämistä ja valintaa tukevien menetelmien käytön, joiden toteuttaminen muilla keinoin olisi käytännössä mahdotonta.

## 5.2 Kontekstin visualisointi

Kyselyn tulosten esittämistapa ei tietenkään vaikuta kyselyjen tietosisältöön tai siihen, millaisia kyselyjä voidaan muodostaa. Jos tiedon tarvitsija olisi tietokone, kyselyjen tulokset voitaisiin esittää vaikkapa binäärisessä muodossa. Ihmisten kyseessä ollessa tulosten esittämistavalla on kuitenkin erittäin suuri merkitys sille, miten suuri osa tuloksen tietosisällöstä on todellisuudessa hyödynnettävissä.

Yksinkertainen ja tehokas tapa helpottaa tiedon tulkintaa on esittää tietoa järjesteltyinä. Sen sijaan, että tiedot esitetään sattumanvaraisessa järjestyksessä, ne esitetään rakenteisesti tai suhteutettuna johonkin. Esimerkkejä tiedon järjestykseen kohdistuvista kyselyistä ovat: *'luokittele ideat niiden tarkoituksen mukaan'* ja *'järjestä yhteisöt sen perusteella, miten paljon niiden julkaisemien dokumenttien sisältämiin ideoihin on viitattu ennen kuin yhteisö on dokumentin julkaissut'*. Näissä kyselyissä resurssit suhteutetaan mitta-asteikkoon. Jonkin resurssin lisääminen tai poistaminen tulosjoukosta ei vaikuta muiden resurssien asemaan asteikolla eikä niiden keskinäiseen järjestykseen. Resursseja voidaan järjestää paitsi suhteessa mitta-asteikkoihin, myös suhteessa muihin resursseihin. Tässä luvussa tarkastellaan tällaisen järjestämisen mahdollisuuksia.

Informaation visualisoinnilla tarkoitetaan tietojenkäsittelyllisiä tekniikoita, joiden avulla dataa voidaan esittää visuaalisesti ja spatiaalisesti [Che02]. Yksi keskeinen informaation visualisoinnin sovelluskohde on esittää resurssien välillä vallitsevia suhteita. Käyttöliittymän teknologiset ominaisuudet määräävät viime kädessä visualisoinnin mahdollisuudet, vrt. keskimääräistä suurempi näyttö [BGS01] ja keskimääräistä pienempi näyttö [BjR00]. Käyttöliittymäteknologioiden merkitystä voi vähentää välttämällä lopullisen visuaalisen ilmeen kiinnittämistä. Tässä tarkastellaankin melko yleisellä tasolla, miten suhteelliseen järjestämiseen perustuvat visualisointitekniikat ja kontekstieto niiden syöteaineistona voivat tukea toisiaan.

Doylen artikkeli 'semanttisista tiekartoista' (semantic road maps) [Doy61] on ensimmäisiä, joissa tarkastellaan visualisoinnin mahdollisuuksia tiedon haun tehostamiseksi. Perusajatus on, että dokumenttien etsiminen on helpompaa, jos dokumenttien ominaisuudet esitetään visuaalisesti nopeasti hahmotettavalla tavalla ja dokumentit järjestetään selailun helpottamiseksi niiden ominaisuuksien mukaan. Doyle visioi ratkaisua, jossa dokumenteista luodaan tietokoneen näyttöruudulle visuaalisia edustajia ja niiden välille yhteyksiä, jolloin tietojen selailua voi suorittaa tietokoneen avulla. Kun dokumenttien ominaisuudet on kuvattu tietokoneelle, dokumenttien edustajia ja niiden välisiä assosiaatioita voidaan luoda dynaamisesti ja automaattisesti tiedon tarvitsijan tarpeiden mukaan.

Pelkästään resurssien ominaisuuksien vertailuun perustuvaa suhteellista järjestämistä kutsutaan ryvästämiseksi (clustering). Lin esittelee erilaisia ryvästämiseen perustuvia visualisointitekniikoita, visualisointiin perustuvan selauksen ja kyselyihin perustuvan etsinnän eroja, sekä esittelee tarkemmin, miten 2-ulotteinen dokumenttikartta voidaan muodostaa dokumentin sisältämien sanojen perusteella ja tiettyä algoritmia (Self-Organizing Map [Koh00]) käyttäen [Lin97]. Toinen hyvä, useita kuvia sisältävä esimerkki ryvästämisen käytöstä löytyy Boyackin, Wyliein ja Davidsonin artikkelista [BWD01]. Ryvästämisen tuloksen ei tarvitse olla kaksi- tai useampiulotteinen, vaan ryppäitä voidaan esittää myös listamuotoisesti, mistä esimerkkinä mainittakoon Scatter/Gather –tekniikka [HeP96].

Ryvästämisen lisäksi toinen perustekniikka resurssien suhteelliseksi järjestämiseksi on verkon piirtäminen (graph drawing). Verkon piirtäminen on mahdollista, kun

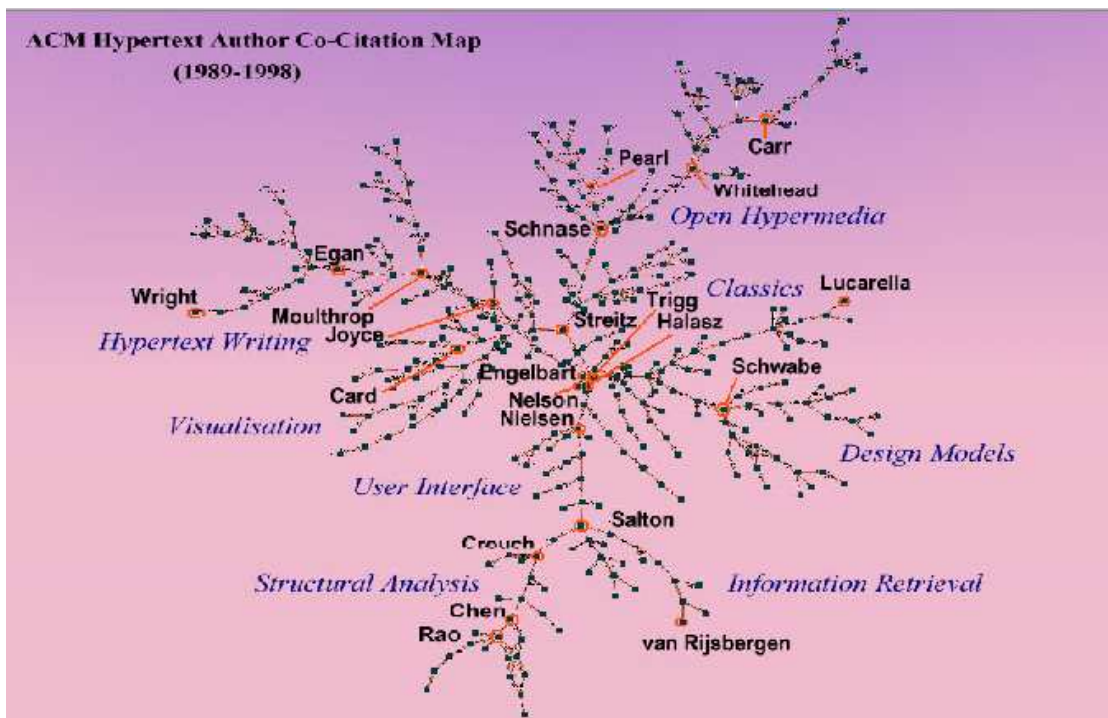
syöteaineisto sisältää suoraan tietoa resurssien välisistä assosiaatioista. Tällöin resurssit voidaan nähdä verkon solmuina ja resurssien välisiä suhteita kuvaavat linkit solmujen välisinä kaarina. Visuaalisesti perusongelmana on solmujen ja niitä yhdistävien käyrien sijoittelu helposti hahmotettavassa muodossa. Herman, Melancon ja Marshall esittelevät verkon piirtämisen pääperiaatteet visualisoinnin näkökulmasta [HMM00].

Verkon piirtämisen tekniikoita käyttäen voidaan muodostaa visuaalisia esityksiä, joita voi käyttää esimerkiksi selaamista helpottavina karttoina [Muk99], tiedonvälitysjärjestelmän käyttäjien, tässä tapauksessa tuottajien ja julkaisijoiden välisten sosiaalisten suhteiden tarkasteluun [Miz94, Fre00] tai tiedon alueiden välisten suhteiden hahmottamiseen [Sma99]. Erittäin hyödyllinen palvelu on saattaa tiedon tarvitsijan tietoon myös henkilöitä, jotka tuntevat tiedon tarpeen kohteena olevan aiheen hyvin. Onhan niin, etteivät automaattiset menetelmät voi korvata sitä apua, jonka asian tunteva henkilö voi ideoiden ymmärtämiseksi ja ongelmien ratkaisemiseksi antaa. Henkilöiden löytämisen menetelmiä esitellään useissa artikkeleissa [Fon97, KSS97, YVS99].

Kuvassa 8 on esimerkki siitä, miltä kontekstin visuaalinen kuva voisi näyttää. Kuva esittää dokumenttien tuottajia ja tuottajien välisiä suhteita. Koska samalla tiedon alueella toimivilla tuottajilla on yleensä tiiveimmät suhteet, kuvaan on myös voitu merkitä, mille alueille (Hypertext Writing, Open Hypermedia jne.) tuottajat lähinnä sijoittuvat. Chen on tuottanut useita vastaavia visualisointeja [ChP01, CPO01]. Kontekstia kuvaava metatieto sisältää tarvittavaa tietoa juuri tämäntyyppisten kuvien muodostamiseksi. Muita visualisointeja, joita voi suoraan pitää sekä hyödyllisinä että mahdollisina toteuttaa, esittävät edellisten lukujen yhteydessä esitetyt kuvat 3 ja 5.

Ryvästämisen ja verkon piirtämisen tekniikkoja voidaan yhdistellä. Kahden resurssin välistä relaatiota voidaan pitää merkkinä siitä, että resursseilla on yhteisiä ominaisuuksia. Relation ei tarvitse olla suora, vaan esimerkiksi yhteinen viitattava tai viittaava resurssi voi olla yhdistävä tekijä [Sma73]. Ryppäiden välille voidaan muodostaa linkkejä käyttämällä hyväksi tietoa ryppäisiin kuuluvien resurssien välisistä linkeistä. Esimerkiksi organisaatioiden välisiä yhteyksiä voidaan mallintaa sen





Kuva 8. Esimerkki kontekstin visualisoinnista [ChC99].

perusteella, miten paljon yhteisiä ideoita ja esityksiä ne ovat julkaisseet, kuinka paljon julkaistuilla ideoilla ja esityksillä on ollut yhteisiä tuottajia ja mitkä yhteisöt ovat myöhemmin julkaisseet organisaatioissa julkaistua materiaalia. Tekniikoita yhdistelemällä voidaan luoda visuaalisia esityksiä, joita voi tarkastella eri tarkkuustasoilla ja 'zoomata' eri näkymien välillä. Small ja Garfield esittelevät eri tarkkuustasojen käyttöä [SmG85]. Visuaalisen näkymän ei tarvitse olla staattinen, vaan sitä voidaan muodostaa sitä mukaa, kun käyttäjä selailee tietoa tai muodostaa uusia kyselyjä ja eri näkymiä voidaan tallentaa myöhempää käyttöä varten [MRC95, BaW97].

Kontekstitieto soveltuu hyvin visualisointitekniikoiden syöteaineistoksi. Resurssien ominaisuudet esitetään standardimuotoisesti resurssista erillisenä metatietona, joten resurssien formaatti tai esitystapa ei vaikuta syöteaineiston saatavuuteen tai sisältöön. Linkit ovat kaksisuuntaisia, joten viittaavien resurssien tarkastelu on teknisesti helppoa. Linkkityyppien käyttö mahdollistaa linkkien merkityksen ja painoarvojen huomioimisen. Eri resurssityypit (ideat, esitykset, arkistot jne.) voidaan erottaa

toisistaan, jolloin tietoa näiden välisistä suhteista voidaan hyödyntää. Julkaisijoiden hierarkia (ml. ideoiden ja esitysten tuottajat) muodostaa luonnollisen lähtökohdan resurssien ryvästämiseksi ja näkymien jakamiselle useisiin tarkkuustasoihin. Tärkein ominaisuus lienee se, että avoimen arkkitehtuurin ansiosta resurssien näkyvyys ja metatietojen saatavuus eivät ole kiinni erillisistä tiedon välittäjistä, kuten viitetietokantojen ylläpitäjistä.

Kontekstin visualisointiin perustuva tiedon hakeminen eli tiedon selaaminen ja perinteisempi yksiulotteisen listan tuottava tiedon haku kyselyn avulla eivät ole toisiaan poissulkevia käyttöliittymäratkaisuja. Tilanne on sama kuin kontekstia kuvaavan tiedon ja sisältöä kuvaavan tiedon käytössä – joihinkin asioihin toinen ratkaisu sopii paremmin, joihinkin toisiin taas toinen. Yhteistä on myös se, että eri tekniikoiden yhdistäminen saattaa usein olla kaikkein tehokkain ratkaisu. Leuskin ja Allanin Lighthouse on yksi esimerkki perinteisen listaperustaisen ja visuaalisen esittämistavan yhdistämiseen perustuvasta käyttöliittymäratkaisusta [LeA00].

### 5.3 Idean arvon mittaaminen

Idean arvo voidaan yrittää todeta ymmärtämällä idea, analysoimalla sitä, käyttämällä sitä ja vertailemalla sitä muihin vaihtoehtoisiin ideoihin. Jo yhden idean arvon määrittäminen tällä tavoin voi vaatia huomattavan paljon resursseja, kuten ymmärtämystä, aikaa ja välineitä. Aiheesta kuin aiheesta on yleensä tuotettu tai tullaan tuottamaan kymmeniä, satoja, jopa tuhansia vaihtoehtoisia tai ainakin vaihtoehtoisilta vaikuttavilta ideoita. Realistisesti ajateltuna tiedon tarvitsijoilla ei ole jokaisessa tiedon tarvetilanteessa mahdollisuuksia tarkasti arvioida jokaista ideaa. Siksi käytettävät ideat valitaan löydettyjen ideoiden joukosta joidenkin heuristiikkojen avulla, joiden toivotaan riittävän hyvin erottelevan tarvittavat ja tarpeettomat ideat toisistaan. Mahdollisuudet ideoiden valintaan riippuvat siitä, mitä lisätietoa ideoista on saatavilla. Tarkastellaan seuraavaksi sitä, millaisia arvioita idean ominaisuuksista voidaan tehdä idean kontekstia kuvaavan metatiedon avulla.

Kun idean tuottajat ovat julkaisseet idean, sitä voidaan arvioida pelkästään idean sisällön ja tuottokontekstin perusteella. Jos ja kun idean tuottanut organisaatio julkaisee idean, voidaan organisaatiota koskevia tietoja ottaa huomioon. Tietenkin myös idean julkaisematta jättäminen voi olla idean arvoa mittaava tekijä. Tämän jälkeen julkaisukontekstin tietoja voidaan hyödyntää sen mukaan, millaiset yhteisöt julkaisevat idean. Kun idea on julkaistu, siihen voidaan alkaa viitata, jolloin myös idean käyttökontekstin hyödyntäminen tulee mahdolliseksi. Arvon mittaamisen kannalta tässä prosessissa on oleellista se, että idean sisällön ja saatavillaoloajan lisäksi idean käyttöön vaikuttavat myös idean tuottajan, idean tuottaneen organisaation ja idean julkaisijoiden ominaisuudet. Jos tuottajan ja organisaation tunnettuus lisäävät idean käyttöä ja idean käyttö niiden tunnettuutta [Mer68] ja ideoiden julkaisemiseen ja käyttöön vaikuttavat niiden arvon lisäksi myös ideoiden tuottajien, julkaisijoiden ja käyttäjien väliset sosiaaliset (insitutionaaliset, poliittiset tms.) kytkökset [Kos98], niin idean todellisen arvon määrittäminen ei voi koskaan olla mikään yksinkertainen laskutoimitus.

Tarkastellaan lähemmin yhden idean arvon kriteerin määrittämistä, nimittäin idean vaikuttavuuden (impact) laskemista ja kontekstiedosta saatavia etuja sen määrittämiseksi. Viittausanalyysin (citation analysis) perusajatuksena on, että mitä enemmän dokumenttiin viitataan, sitä enemmän vaikutusta dokumentilla on ollut. Ainakin seuraaviin viittausanalyysin ongelmiin kontekstiedon olemassaolo tarjoaa ratkaisuja (kohdat 1-4 ovat lähteestä [Gar79] ja kohdat 5-8 lähteestä [MaM96]):

1) Ongelma: viittausanalyysissä ei erotella positiivista ja negatiivista palautetta, vaan kaikkia viittauksia pidetään merkinä positiivisesta vaikutuksesta.

Ratkaisu: viittauksen tyyppi tiedetään (perustuu, tukee, horjuttaa, jne.), joten päästään vielä pelkkää positiivinen/negatiivinen –jakoakin tarkempaan erotteluun.

2) Ongelma: viittauskäytännöt vaihtelevat idean tyyppin mukaan. Menetelmäideat, kuten mittausmenetelmät ja laskentamenetelmät (l. miten-ideat ja tulokset) keräävät eniten viittauksia, koska useimmat ideoista perustuvat juuri tällaisiin ideoihin.

Ratkaisu: idean tyyppi tiedetään, joten se voidaan ottaa huomioon sekä tarkasteltavan että viittaavien ideoiden kohdalla.

3) Ongelma: itseensä viittaaminen (self-citation). Käytön määrää voidaan 'keinotekoisesti' lisätä viittamalla erityisesti omiin ideoihin, omia ideoita tukevien tuottajien ideoihin ja ideoihin organisaatioiden ja organisaatioblokkien sisällä.

Ratkaisu: koska tuottajat ja julkaisijat kuuluvat malliin, 'viittausryppäät' voidaan automaattisesti tunnistaa ja niiden sisäisten viittausten merkitystä vähentää.

4) Ongelma: julkaisukäytännöt ovat teknisesti kehittymättömiä. Usein tuottajista on lueteltu vain ensimmäinen tai vain osa, saman nimisiä tuottajia ei pystytä erottamaan, lähdeluetteloissa on yllättävän paljon virheitä, joiden seurauksena viittauksia ei tunnisteta, viittausten kokoajat pystyvät käsittelemään vain murto-osan kaikista julkaisuista jne.

Ratkaisu: standardoitu metatiedon ja tunnisteiden käyttö, kaksisuuntaisten linkkien käyttö ja tietojenkäsittelyllisten menetelmien käyttö poistavat tehokkaasti tämäntyyppisiä ongelmia.

5) Ongelma: kaikkiin tärkeisiin vaikutteisiin ja vaikuttajiin ei viitata.

Ratkaisu: kun viittauslista on erotettu esityksestä, voidaan taustoittaa-linkin avulla luetella myös lähteitä, jotka eivät retoriikan takia mahdu tai sovi tekstissä mainittaviksi.

6) Ongelma: viittaukset painottuvat epätasaisesti. Viittaajat viittaavat niihin lähteisiin, jotka he ovat löytäneet, eivät alkuperäisiin lähteisiin. Alkuperäisten artikkelien sijaan viitataan esimerkiksi koosteisiin.

Ratkaisu: tuottokontekstin linkkipolkuja seuraamalla alkuperäiset lähteet ovat helposti löydettävissä.

7) Ongelma: huono saatavuus ja näkyvyys rajoittavat dokumenttien käyttöä.

Ratkaisu: julkaisijan valinta ei kiinnitä julkaisukanavaa (kun julkaisijalla on tekijänoikeudet vain dokumenttiin, ei ideaan tai esitykseen), joten saatavuutta ja näkyvyyttä voidaan lisätä julkaisemalla idea monessa paikassa ja usealla julkaisutasolla.

8) Ongelma: vaikka idean olemassaolo tiedettäisiin, niin jos dokumenttia ei löydetä, ei ideaa voida käyttää.

Ratkaisu: jokaisessa idean metatietokuvauksessa on tiedot kaikista idean sisältävistä dokumenteista ja linkit tietovarastoihin, joista dokumentti on saatavissa.

Merkityksettömyys - vaikuttavuus on tietenkin vain yksi lukemattomista arvon kannalta kiinnostavista ominaisuuksista, joita ideoista voidaan pyrkiä mittaamaan. Muita kriteerejä voisivat olla esimerkiksi tuntemattomuus - tunnettuus, luotettavuus - potentiaalisuus ja vakaus - vakiintumattomuus. Otetaan esimerkkinä ominaisuus tavanomaisuus - poikkeuksellisuus.

Poikkeuksellisuutta voisivat heijastaa seuraavat idean ominaisuudet:

- julkaisija ei ole ennen julkaissut muita idean tuottajien tuottamia ideoita
- julkaisija ei ole ennen julkaissut aiempien idean julkaisijoiden julkaisemia ideoita
- idean tuottaja on julkaissut aikaisemmat ideansa muiden julkaisijoiden välityksellä
- ideasta on heti ensimmäisen julkaisun jälkeen saatavilla monta eri esitystä
- idea on ensimmäinen, joka perustuu johonkin ideaan
- idea on ensimmäinen, joka käsittelee (tarkastelee, tukee tai horjuttaa) jotain ideaa
- idea horjuttaa aikaisemmin vain tukea saanutta ideaa tai päinvastoin
- idean lähteet ovat hyvin laajalta alueelta
- ideaan perustuu monia tarkoitukseltaan eri tyyppisiä ideoita
- ideaan viittaa paljon sekä sitä tukevia että sitä horjuttavia ideoita
- idean julkaisu tasolta toiselle on tapahtunut poikkeuksellisen hitaasti tai nopeasti

Nämä ovat tietenkin pelkkiä arvauksia siitä, mitkä ominaisuudet voisivat korreloida poikkeuksellisuuden kanssa. Enempää ne eivät voikaan olla, koska ominaisuuksien merkitys voidaan todeta vain tarkastelemalla tiedonvälitysjärjestelmän käyttöä aidossa ympäristössä. Myös esimerkit jatkossa ovat samalla tavalla vain suuntaa-antavia.

Jos ideoiden arvoon vaikuttavia ominaisuuksia voidaan mitata, niin niiden perusteella voidaan mitata myös ideoiden tuottajien (ml. organisaatiot) ja ideoita julkaisevien yhteisöjen ominaisuuksia. Näiden mittaamisesta on hyötyä ainakin kahdesta syystä. Ensinnäkin tuottajien, organisaatioiden ja yhteisöjen merkitys idean ominaisuuksiin voidaan ottaa huomioon. Esimerkiksi jos kahta ideaa on käytetty yhtä paljon, voidaan sitä ideaa pitää potentiaalisesti arvokkaampana, jonka tuottajat ja julkaisijat ovat tuntemattomampia, koska jo tunnettuus sinänsä lisää käytön todennäköisyyttä.

Toiseksi idean tuottajien ja julkaisijoiden ominaisuudet saattavat korreloida idean arvon kanssa. Esimerkiksi jos ideoiden tuottaja on tietyn aihealueen asiantuntija, voi arvata, että tuottajan tuottamat, kyseistä aihetta käsittelevät ideat eivät oletusarvoisesti sisällä täysin virheellistä tietoa. Vastaavasti jos tiedetään, että jokin yhteisö arvioi julkaisemansa ideat erittäin tarkasti, voidaan idean julkaisemista tämän yhteisön toimesta pitää merkityksellisempänä kuin ehkä jonkin toisen yhteisön toimesta. Otetaan tuottajien ominaisuuksien mittaamisesta esimerkkinä ominaisuus spesiaalisuus - laaja-alaisuus:

Spesiaalialan tietämystä voisivat heijastaa seuraavat ominaisuudet:

- tuotetut ideat ovat tarkoitukseltaan samantyyppisiä
- tuotettujen ideoiden lähteet ovat pitkälti samoja
- tuotettujen ideoiden lähteillä on pitkälti samoja julkaisijoita
- ideoita käyttää tietty joukko
- ideoita julkaisevat samat julkaisijat
- tuottaja tuottaa ideoita yhteistyössä muiden kanssa
- tuottaja tuottaa paljon tuloksia tai analyysejä

Laaja-alaisuutta voisivat heijastaa seuraavat ominaisuudet:

- tuotetuilla ideoilla on monipuolisesti erilaisia lähteitä
- tuotetuilla ideoilla on lukuisia eri yhteisöjulkaisijoita
- ideoita julkaistaan korkeilla julkaisutasoilla
- tuottaja tuottaa paljon miksi-ideoita, mitä-ideoita tai koosteita
- tuottaja tuottaa ideoita myös yksin
- tuotettujen ideoiden käyttäjät ovat laajalta alueelta
- tuotetuista ideoista on saatavilla monta erilaista esitystä

Esimerkkinä yhteisöjen ominaisuuksien mittaamisesta voisi olla ominaisuus ideoiden arvioija - ideoiden kokoaja.

Ideoita arvioivia yhteisöjä voisivat kuvata seuraavat ominaisuudet:

- yhteisö julkaisee lähinnä uusia ideoita
- yhteisö julkaisee ideoita harvoin
- alemman tason julkaisun ja yhteisön julkaisun välillä on viivettä

- yhteisön julkaisemia ideoita käytetään enemmän kuin muita idean tuottajien ja idean julkaisseen organisaation ideoita
- julkaistuilla ideoilla on paljon yhteisiä lähteitä
- yhteisö julkaisee vain pienen osan niistä ideoista, joita yhteisön julkaisemien ideoiden tuottajat ja organisaatiot tuottavat
- suuri osa yhteisön julkaisemista ideoista julkaistaan myös ylemmillä tasoilla
- yhteisön julkaisemia ideoita horjutetaan vähän
- yhteisö ei yleensä julkaise koosteita

Koska idean arvon käsite on hyvin epämääräinen ja subjektiivinen, tiedon tarvitsijan tulee tilannekohtaisesti voida määritellä, mitkä ovat arvokkaiden ideoiden ominaisuuksia. Joissain tapauksissa tiedon tarpeen kohteena voivat olla esimerkiksi 'tuntemattomat, potentiaaliset, vakiintumattomat, yleisneron tuottamat ja ideoita arvioivan yhteisön julkaisemat ideat' yhtä hyvin kuin joissain toisissa tilanteissa 'tunnetut, luotettavat, vakaat, specialistin tuottamat ja ideoita kokoavien yhteisöjen julkaisemat' ideat. Lisäksi idean tarvitsijalla tulee olla mahdollisuus vaihtaa, lisätä, poistaa ja muuttaa mittareita sitä mukaa, kun uusia mittareita kehitetään ja olemassaolevia säädetään.

Käyttäjät tuskin itse osaavat määritellä ja koodata arvon mittareita. Siksi saatetaan tarvita erityisiä 'arvoteknikkoja' tai 'bibliometriikka-insinöörejä' (vrt. Bushin 'trailblazers' [Bus45]), jotka tuntevat tiedonvälityksen käytäntöjä ja osaavat muuttaa arvon eri määritelmiä metatietoa käsitteleviksi laskutoimituksiksi. Näiden vastuulle jää sellaisten mittareiden kehittäminen, jotka tehokkaasti mittaavat väitetyjä ideoiden ominaisuuksia ja tiedon tarvitsijan vastuulle jää mittareiden valinta ja järkevä käyttö.

#### 5.4 Automaattinen tiedon valinta

Yhtenä teoreettisena tavoitteena tiedonvälitysjärjestelmiä kehitettäessä voi pitää järjestelmää, jossa tarvittavat tiedot toimitetaan automaattisesti tiedon tarvitsijalle tiedon tarpeiden ja saatavilla olevien tietojen muuttuessa. Tämän tavoitteen

saavuttamiseksi tarvitaan menetelmiä, joiden avulla tietoa voidaan valita käyttäjän puolesta, ts. automaattisesti.

Automaattisesti tietoa valikoivia menetelmiä on kehitetty etenkin viestintäjärjestelmiin, joista esimerkkejä ovat sähköposti ja keskusteluryhmät. Viestintäjärjestelmissä kynnyks uusien viestien tuottoon on matala ja tiedon tuottaja voi lähettää ('työntää') vastaanottajille tietoa ilman, että nämä ovat sitä aktiivisesti hakeneet. Seurauksena on helposti tietotulva, jossa hyödylliset viestit hukkuvat hyödyttömien joukkoon. Automaattisesti tietoa valitsevien menetelmien tarkoituksena on tällöin karsia hyödyttömät viestit pois tietovirrasta eli estää käyttäjää näkemästä osaa viesteistä. Viestien välityksen näkökulmasta automaattisen tiedon valinnan perusteita esittelevät Hiltz ja Turoff [HiT85] sekä Malone, Grant ja Turbak [MGT86]. Tiedon karsinnalle vastakkainen tapa valita tietoa käyttäjän puolesta on esittää käyttäjälle tietoa, jota käyttäjä ei olisi muuten välttämättä löytänyt tai tullut valinneeaksi. Tällaisia tietoa käyttäjille työntäviä menetelmiä on alettu kutsua suosittelujärjestelmiksi (recommendation systems) [ReV97]. Niille on löytynyt käyttökohteita muun muassa sähköisillä kauppapaikoilla, joilla ostajakandidaatille pyritään automaattisesti suosittelemaan ostajaa toivottavasti viihdyttävää materiaalia, kuten kirjoja, elokuvia tai musiikkia [SKR99].

Käyttäjän näkökulmasta automaattinen tiedon valinta on joko automaattista tiedon karsintaa tai automaattista tiedon suosittelua. Näkökulmasta riippumatta lopullista valintavaihetta tietojenkäsittelyllisenä menetelmänä kutsutaan tiedon suodattamiseksi (information filtering). Tiedon suodattaminen tarkoittaa sitä, että staattista tiedon tarpeen kuvausta eli käyttäjäprofiilia verrataan dynaamiseen tietovirtaan, erotuksena perinteisestä tiedon hausta (information retrieval), jossa dynaamisesti muuttuvia tiedon tarpeen kuvauksia (kyselyjä) verrataan staattiseen tietoaaineistoon tietovarasto(i)ssa [BeC92].

Suodatustekniikat eivät määrittele sitä, miten tietovirta luodaan. Kun tietovirtaa ei valmiiksi ole, on tiedot valikoinnin lisäksi myös löydettävä. Tällöin tarvitaan menetelmiä, jotka hakevat tietoa tietovarastoista automaattisesti. 'Puoliautomaattisesti' tämä onnistuu siten, että käyttäjä eksplisiittisesti määrittelee, millaista tietoa hän haluaa itselleen suositeltavan. Aidosti automaattinen tiedon etsiminen merkitsee sitä,



että automaattisen valinnan lisäksi myös päätelmiä tiedon tarpeesta tehdään automaattisesti. Maes kutsuu tällaista tiedon tarvetta arvioivaa ohjelmistoa (henkilökohtaiseksi digitaaliseksi) assistentiksi [Mae94]. Kokonaisuutena automaattinen tiedon valinta koostuu siis automaattisesta tiedon tarpeen kartoittamisesta (assistenttiohjelmistot) ja automaattisesta tiedon arvioinnista (tiedon suodatus).

Tiedon suodatuksen toimivuus riippuu siitä, miten hyvin käyttäjän kuvaus eli profiili ja valittavan aineiston kuvaus keskenään vertautuvat. Perinteisesti aineiston kuvaukset ovat olleet sisällön kuvauksia, yleisimmin asiasanojen ja automaattisesti luotujen indeksien avulla. Yksinkertaisimmillaan profiili voidaan tällöin muodostaa siten, että käyttäjä luettelee sanoja, joiden perusteella hän haluaa resursseja joko suositeltavan tai karsittavan.

Perusratkaisuun on kehitetty useita laajennuksia, jotka perustuvat aineiston kuvauksen monipuolistamiseen, profiilin muodostamiseen käyttäjän puolesta tai näiden kombinaatioihin. On kehitetty menetelmiä, joiden avulla käyttäjien toimintaa, esimerkiksi jonkin dokumentin käyttötiheyttä [OaK98] ja käyttäjän kontekstia (sijaintia, ympäristöä, ajankohtaa, tehtävää, jne.) [ByC01, LaN92, SAW94] voidaan automaattisesti tarkkailla ja kuvata. Lisäksi on olemassa erilaisia agenttiperustaisia ratkaisuja, kuten tiedonhakutilanteita tarkkailevia ja analysoivia ohjelmia [BuH99, RhM00, LFW01]. Yhteistoiminnallisessa suodatuksessa (collaborative filtering) [Gol92] taas käyttäjäprofiilin tietoja jaetaan useiden profiilien kesken.

Kaikkia sisällön kuvaukseen perustuvia suodatusmenetelmiä vaivaa kuitenkin perusongelma, jota erilaiset laajennukset eivät ratkaise: käyttäjälle voidaan suositella resursseja vain sisällön samankaltaisuuden perustuen, koska sisältö on ainut resursseja yhdistävä tekijä. Käyttäjälle voidaan suositella tarkasti ja osuvasti vain sisällöllisesti tiettyihin ryhmiin kuuluvia tai, yhteistoiminnallisen suodatuksen menetelmiä käytettäessä, tiettyjen käyttäjäryhmien arvostamia resursseja.

Tilanteen ongelmallisuus riippuu siitä, millaista tietoa suodatetaan. Viihdemateriaalin, esimerkiksi musiikin, välityksessä tuskin haittaa, että käyttäjälle suositellaan vain sen aihealueen 'tietoa', josta käyttäjä on ennen eniten pitänyt. Sen sijaan esimerkiksi

ideoiden välityksessä ei voi pitää riittävänä sellaista ratkaisua, joka supistaa käytettävän tiedon kenttää jatkuvasti vain spesifimpiin aihealueisiin ja vähentää tiedon välittymistä erilaisten ja poikkeavien tiedon alueiden välillä. Suodatuksen seurauksena tunnettuuden ja käytön välille syntyy helposti positiivisia palautekytkentöjä. Viihdemateriaalin kohdalla nämä eivät suuremmin haittaa, kunhan käyttäjät ovat tyytyväisiä, mutta ideoiden välityksen kyseessä ollessa ne ovat haitallisia. Ideoiden välityksen kohdalla on myös kyseenalaista, voiko sisältöön perustuvan erottelun tarkkuus koskaan olla riittävä sellaisessa tiedonvälitysjärjestelmässä, jossa miltei aiheesta kuin aiheesta on saatavilla paljon ideoita.

Sisällön kuvaukseen perustuviin suodatusmenetelmiin verrattuna kontekstitiedon olemassaolo tarjoaa kaksi eittämätöntä etua: resurssien välisten assosiaatioiden kautta yhden resurssin arvosta voidaan tarkemmin tehdä muiden resurssien arvoon vaikuttavia johtopäätöksiä ja kuvaustiedon monipuolisuuden ansiosta valintaa voidaan suorittaa monipuolisemmin. Tuottajien, julkaisijoiden ja käyttäjien välisiä suhteita koskevan tiedon ansiosta tiedon siirtymistä eri tiedonalueiden ja näkemysten välillä voidaan edistää sen sijaan, että tietoa pyrittäisiin eristämään rajattujen sisältöalueiden sisään. Eduksi voidaan laskea myös se, että kontekstitiedon kautta voidaan automaattisen tiedon valinnan kohteeksi ottaa esitysten lisäksi myös kontekstin malliin kuuluvia muita resursseja, kuten ideoiden tuottajia tai julkaisijoita. Tällaisten resurssien karsinta ja suosittelu tai niiden avulla tapahtuva esitysten suositteluhan ei pelkästään esitysten sisältöön perustuen voi onnistua.

Hyvä esimerkki kontekstitiedon hyödyllisyydestä on tilanne, jossa pyritään suosittelemaan tiettyyn ideaan liittyen muita relevanteja ideoita. Budzik toteaa, että usein käyttäjien kannalta hyödyllisiä olisivat ideat, jotka kyseenalaistavat, kritisoivat ja arvioivat kyseessä olevaa ideaa, mutta esityksien samankaltaisuuden vertailuun perustuvien menetelmien avulla tällaisia ideoita on hyvin vaikea automaattisesti valita [Bud00]. Kontekstitiedon avulla käyttäjälle voidaan triviaalisti suositella niitä ideoita, jotka horjuttavat tai ainakin tarkastelevat kyseessä olevaa ideaa. Helposti voidaan kehittää monipuolisempiakin valintasääntöjä. Tässä tapauksessa potentiaalisesti hyödyllisiä voisivat olla esimerkiksi ideaa perustavia ideoita horjuttavat ideat, ideaan perustuvia tai ideaa tukevia ideoita horjuttavat ideat, sellaisten tuottajien tuottamat

ideat, jotka ovat aikaisemmin horjuttaneet idean tuottajan tuottamia ideoita, ja niin edelleen.

Otetaan lopuksi kolme konkreettisempaa esimerkkiä kontekstitiedon hyödyntämisestä automaattisen tiedon valinnan syöteaineistona.

*Tiedon karsinta.* Koska tiedon tuottajilla ja julkaisijoilla ei esitetystä mallissa ole mahdollisuutta lähettää tietoa suoraan tiedon tarvitsijoille, ei viestintäjärjestelmiä vaivaava tietotulva ole mahdollinen. Tietotulva voi kuitenkin syntyä siten, että tiedon tarvitsija tietoa etsiessään löytää liikaa potentiaalisesti hyödyllistä tietoa. Tällöin olisi hyödyllistä pystyä automaattisesti karsimaan 'tietovirrasta' varmasti hyödyttömät ideat. Karsinta voi olla esimerkiksi karsittujen resurssien täydellistä piilottamista, karsittujen resurssien merkitsemistä (l. ehdottamista) karsituiksi tai karsittujen resurssien siirtämistä kaikissa arvioissa, järjestyksissä ja muissa vastaavissa listoissa viimeisiksi. Esitysten sisältöön perustuen, esimerkiksi sanalistojen avulla voidaan kyllä karsia kokonaisia aihealueita, mutta aihealueiden sisällä hyödyttömien ideoiden erottaminen sisällön perusteella on hyvin vaikeaa. Kontekstitietoon perustuva karsinta onkin tehokkaimmillaan aihealueiden sisällä tapahtuvassa hienojakoisemmassa karsinnassa, koska se tarjoaa käyttäjille välineet ilmoittaa, millaista tietoa he pitävät hyödyttömänä.

Esimerkkejä tuottokontekstin hyödyntämiseen perustuvasta karsinnasta ovat: *'karsi kaikki ideat, jotka perustuvat ideaan x'* ja *'karsi kaikki ideat, joiden tuottamiseen tuottaja x on osallistunut'*. Käyttökontekstin avulla ideoita voidaan karsia vaikkapa seuraavasti: *'karsi ne ideat, joita tuetaan ideaan x perustuvien analyysien toimesta'*. Julkaisukontekstin avulla karsinta on erittäin tehokasta: *'karsi kaikki organisaation x julkaisemat ideat'*, tai peräti *'karsi kaikki ne ideat, joiden tuottajien yksikin idea on julkaistu yhteisön x toimesta'*. Jo mahdollisuus siihen, että tiedon tarvitsijat voivat helposti karsia kokonaisia tuottaja- tai julkaisijaryhmiä, vähentänee tehokkaasti tuottajien ja julkaisijoiden motivaatiota julkaista järjestelmässä todella surkeita ideoita tai tahallisesti vääristellä kontekstitietoja.

*Tiedon tilaaminen.* Suositteleminen on yksinkertaisinta ja vähiten häiritsevää, kun käyttäjälle suositellaan vain materiaalia, jonka käyttäjä on eksplisiittisesti tilannut. Tilaaminen sopii luontevasti uusien resurssien seurantaan, koska uusien resurssien ilmestyminen

järjestelmään luo tarvittavan tietovirran, johon suodatusmenetelmiä voidaan kohdistaa. Esimerkki tiedon tilaamisesta, joka on henkilökohtaisten metatietovarastojen kautta helppo toteuttaa ja joka on varmasti erittäin hyödyllinen palvelu kaikille järjestelmän käyttäjille, on *'ilmoita kaikista uusista ideoista, jotka käyttävät yhtenä lähteenä jotakin tuottamaani ideaa'*. Kontekstietietoon perustuen tietoa voidaan tilata lukemattomilla muillakin perusteilla: *'ilmoita kaikista ideaa x horjuttavista ideoista'*, *'ilmoita kaikista niistä organisaation x julkaisemista ideoista, jotka julkaistaan jonkin yhteisön toimesta'*, *'ilmoita, kun ideasta x ilmestyy suomen- tai englanninkielinen esitys'* ja niin edelleen.

*Automaattinen tuottajien ja julkaisijoiden etsintä.* Hajautetun arkkitehtuurin ongelmana on, että ennen kuin voidaan etsiä tietoa, on löydettävä tietovarastot, joista tietoa etsitään. Siksi sellaista automaattista palvelua voi pitää hyödyllisenä, joka käyttäjän puolesta kerää tietoa tietovarastoista ja suosittelee käyttäjälle tietovarastoja, joista tietoa kannattaisi hakea.

Automaattinen tuottajien ja julkaisijoiden etsintä voidaan toteuttaa siten, että haetaan jokaisesta havaitusta tietovarastosta tiedot kaikista niistä tietovarastoista, joihin tietovarastosta on linkkejä. Iteratiivisesti etenemällä etsintäaluetta voidaan laajentaa. Erilaisia keinoja selaamisen tehostamiseksi on useita [Jos02, PaM02, SYV01].

Ottamalla julkaisijoiden tyypit ja linkkien tyypit huomioon voidaan etsintää mielekkäästi suunnata ja rajata tarpeen mukaan. Etsintä voidaan myös rajoittaa ennestään tuttujen tietovarastojen (ts. tietovarastojen, joista käyttäjä on aikaisemmin etsinyt tietoa) läheisyyteen, jolloin tietämys uusista tietovarastoista laajenee käyttäjän intressien mukaan, eikä tiedon keruu operaationa muodostu liian raskaaksi. Tarvittaessa automaattinen suosittelu voidaan myös rajoittaa koskemaan lähinnä korkean tason yhteisöjä, joiden kautta käyttäjän on itsekin helppo selata alempien tasojen tarjontaa (vrt. luvut 4.3 ja 5.2). Kun käyttäjä sitten etsii tietoa, voidaan automaattisesti kerätyn tietovarastolistan perusteella suositella tiedon tarpeen kannalta relevantteja tuottajia ja julkaisijoita, joiden kautta tietoa kannattaa ehkä hakea.

## 6 Yhteenveto

Mitä enemmän tietoa on saatavilla, sitä enemmän resursseja tarvitaan arvokkaimpien tietojen erottamiseksi muusta tietomassasta. Siksi saatavilla olevan tiedon määrän kasvaessa erilaiset tiedon valintaa helpottavat ratkaisut tulevat yhä tärkeämmiksi.

Tiedon valintaa helpotetaan kuvaamalla tiedot, jolloin valinta voidaan suorittaa kuvauksen perusteella. Yleisimmin kuvataan tiedon sisältöä. Itse sisällön lisäksi tiedon arvoa voi päätellä monesta muustakin seikasta, kuten siitä, mistä näkökulmasta tieto on tuotettu, mihin tietoihin tieto perustuu, miten ymmärrettävästi tieto on esitetty, mihin tietoa on käytetty tai miten merkittävänä tietoa yleisesti pidetään. Ideat ovat tällaista tietoa, jonka arvoa ei voi käytännössä määritellä pelkästään tiedon sisällön perusteella. Idean arvo riippuu aina myös tiedon tarvitsijasta, idean suhteista muihin ideoihin ja ideoiden toimivuudesta käytännössä. Vaikka idean sisältö ei muutu, idean arvo voi hyvinkin muuttua sitä mukaa, kun ideaa ja idean käsittelemää aihetta koskeva tieto muuttuu.

Ideoiden valintaa voidaan helpottaa kuvaamalla sisällön lisäksi myös niiden suhdetta ympäristöönsä eli idean kontekstiin. Tietojenkäsittelyllisten menetelmien käyttö tekee kontekstitiedon hallinnan käytännössä mahdolliseksi. Internet tarjoaa avoimen tietoteknisen infrastruktuurin, johon perustuen voidaan toteuttaa avoimia ideoidenvälitysjärjestelmiä. Metatiedon avulla järjestelmässä käsiteltäviä tietoresursseja voidaan kuvata itse resursseista riippumattomasti. Metatietoon voidaan kohdistaa automaattisia päättelymenetelmiä, joiden avulla metatiedon tietosisältöä saadaan tehokkaasti hyödynnettyä. Hypermediateknologia mahdollistaa resurssien välisten suhteiden ja assosiaatioiden hallinnan ja kaksisuuntaisten linkkien käyttö resurssien kuvausten dynaamisen muuttamisen ja muuttumisen niiden kontekstin muuttuessa. Käyttöliittymien avulla kontekstitiedon hyödyntämistä voidaan helpottaa.

Ideoidenvälitysjärjestelmän tehokkuuden ratkaisee metatietokuvauksen perustana käytettävä malli, joka määrittelee idean kontekstia kuvaavan metatiedon tietosisällön. Metatietomallin tulisi hyödyntää tietojenkäsittelyllisten menetelmien tuomia

mahdollisuuksia, kuten dynaamisuutta, assosiaatioiden luomista ja automaattisten päättelymenetelmien käyttöä, mutta ennen kaikkea sen tulee olla toimiva kompromissi monipuolisuuden eli kontekstitiedon ilmaisuvoiman ja monimutkaisuuden eli kontekstitiedon luomisen ja hyödyntämisen vaikeuden välillä.

Näihin tavoitteisiin perustuen tutkielmassa esitellään idean kontekstia kuvaavan metatiedon malli. Kontekstin kuvaus jakautuu kolmeen osaan: tuottokontekstiin, julkaisukontekstiin ja käyttökontekstiin. Idean tuottokontekstiin kuuluvat idean tuottoa kuvaavat tiedot, julkaisukontekstiin kuuluvat idean esityksiä ja julkaisuja kuvaavat tiedot ja käyttökontekstiin kuuluvat idean käyttöä koskevat tiedot.

Idean tuottaja kuvaa idean ennen kaikkea ilmoittamalla idean tuottajat, tuottoajankohdan, idean tarkoituksen (miksi-idea, mitä-idea, miten-idea, tulos, analyysi, raportti, kooste tai erikseen luokittelematon idea) ja idean tuotossa käytetyt tärkeimmät lähteet käyttötarkoituksineen (perustava, taustoittava, tarkasteltava, tuettu tai horjutettu idea). Esityksen tuottaja kuvaa esityksen erityisesti ilmoittamalla esityksen tuottajat, tuottoajankohdan ja esityksen sisältämät ideat. Julkaisija kuvaa julkaisun ennen kaikkea ilmoittamalla julkaisijan tyypin (idean tai esityksen tuottaja, organisaatio, yhteisö tai arkisto), julkaisuajankohdan, julkaisuun mahdollisesti liittyvät viestintäkanavat, julkaisun sisältämät esitykset sekä tietenkin dokumentin, johon julkaisutiedot liittyvät. Idean käyttäjä kuvaa sen käyttöä joko antamalla ideaa koskevaa palautetta viestintäkanavien kautta tai viittaamalla ideaan uusissa tuottamisissa ideoissa, jolloin käyttötarkoitus nähdään viittauksen tyypin perusteella (perustuva, taustoittuva, tarkasteleva, tukeva tai horjuttava idea).

Esitettyyn kontekstin kuvaukseen perustuen ideoiden ja muiden resurssityyppien välille muodostuu dynaamisesti muuttuva assosiaatioiden ja suhteiden verkko, joka mahdollistaa tehokkaiden ideoiden merkitystä ja arvoa koskevien kyselyjen ja päättelyjen tekemisen. Kontekstin kuvauksen monipuolisuudesta johtuen tiedon tarvitsijat voivat hyödyntää kontekstitietoja tehokkaasti vain silloin, kun metatietokuvaukseen perustuen voidaan toteuttaa käyttöliittymäratkaisuja, jotka abstrahoivat kontekstitiedon käsittelyn tiedon tarvitsijan kannalta riittävän yksinkertaiselle tasolle. Tutkielmassa kontekstitiedon käyttöä tarkasteltiin kontekstin

visualisoinnin, ideoiden arvon mittaamisen ja automaattisen tiedon valinnan näkökulmista.

Kontekstia visualisoimalla helpotetaan resurssien etsimistä, selaamista, käsittämistä, hahmottamista ja ymmärtämistä sekä resurssien välisten suhteiden ja assosiaatioiden tarkastelua ja analyysiä. Kontekstitieto soveltuu hyvin visualisointitekniikoiden syöteaineistoksi, koska se sisältää runsaasti resurssien välisiä suhteita kuvaavaa tietoa. Linkkityyppien käyttö mahdollistaa linkkien merkityksen ja painoarvojen huomioimisen. Eri resurssityypit (ideat, esitykset, arkistot jne.) voidaan erottaa toisistaan, jolloin tietoa näiden välisistä suhteista voidaan hyödyntää. Julkaisijoiden hierarkia muodostaa luonnollisen lähtökohdan resurssien ryvästämiseksi ja visuaalisten näkymien kuten tiedon alueita kuvaavien karttojen jakamiselle useisiin tarkkuustasoihin. Avoimen arkkitehtuurin ansiosta resurssien näkyvyys ja metatietojen saatavuus eivät ole kiinni erillisistä tiedon välittäjistä kuten viitetietokantojen ylläpitäjistä.

Metatiedon perusteella resurssien ominaisuuksia voidaan automaattisesti mitata ja tehdä tästä resurssin arvoa koskevia päätelmiä. Kontekstitiedon olemassaolo helpottaa esimerkiksi vaikuttavuuden (impact) määrittämisestä muun muassa siksi, että viitattavien ja viittaavien ideoiden sekä viittausten tyypit voidaan ottaa huomioon ja ideoiden tuottajien ja julkaisijoiden keskinäisiä ja välisiä suhteita päästään tarkastelemaan. Syöteaineiston ilmaisuvoiman ansiosta voidaan merkityksettömyys-vaikuttavuus -akselin lisäksi pyrkiä mittaamaan monia muitakin arvon kannalta kiinnostavia ominaisuuksia, joita voisivat olla esimerkiksi tuntemattomuus - tunnettuus, luotettavuus - potentiaalisuus, vakaus - vakiintumattomuus, tavanomaisuus - poikkeuksellisuus, spesiaalisuus - laaja-alaisuus ja esimerkkinä julkaisijoihin kohdistuvasta arvioinnista ominaisuus ideoiden arvioija - ideoiden kokoaja. Avoimessa arkkitehtuurissa tiedon tarvitsijalle voidaan jättää mahdollisuus valita, vaihtaa, lisätä, poistaa ja muuttaa mittareita, kun uusia mittareita kehitetään ja olemassaolevia säädetään.

Automaattinen tiedon valinta perustuu tiedon suodatuksen menetelmien ja tiedon tarvetta profiloivien digitaalisten assistenttien käyttöön. Kontekstitieto on tehokasta syöteaineistoa tällaisille menetelmille kahdesta syystä: resurssien välisten

assosiaatioiden kautta resurssien arvoista voidaan tarkemmin tehdä muiden resurssien arvoa koskevia johtopäätöksiä ja kuvaustiedon monipuolisuuden ansiosta valintaa voidaan suorittaa pelkästään sisältöön perustuvia menetelmiä monipuolisemmin. Tuottajien, julkaisijoiden ja käyttäjien välisiä suhteita koskevan tiedon ansiosta tiedon siirtymistä eri tiedonalueiden ja näkemysten välillä voidaan edistää sen sijaan, että tietoa eristettäisiin rajattujen sisältöalueiden sisään. Eduksi voidaan laskea myös se, että kontekstitiedon avulla voidaan automaattisen tiedon valinnan kohteeksi ottaa esitysten lisäksi myös kontekstin malliin kuuluvia muita resursseja, kuten ideoiden tuottajia tai julkaisijoita.

Kaiken kaikkiaan tutkielma osoittaa, että ideoiden välittymistä ja valintaa olisi ainakin periaatteessa mahdollista tehostaa huomattavastikin tietojenkäsittelyllisiä menetelmiä ja niiden mahdollisuudet huomioon ottavia tiedon välittymisen malleja ja käytäntöjä kehittämällä ja hyödyntämällä.



## Lähteet

[Abe98] Abecker, A., et al., Toward a Technology for Organizational Memories. IEEE Intelligent Systems, May/June, 1998, 40-48.

[Akm01] Akman, V., et al., (toim.), Modeling and Using Context. Proceedings of the Third International and Interdisciplinary Conference, CONTEXT 2001, Dundee, UK, July 27-30, 2001.

[AkS96] Akman, V., Surav, M., Steps Toward Formalizing Context. AI Magazine 17, 3, (1996), 55-72.

[ATH00] Amento, B., Terveen, L., Hill, W., Does "Authority" Mean Quality? Predicting Expert Quality Ratings of Web Documents. Teoksessa Proceedings of the 23rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2000, 296-303.

[BaB97] Bannon, L., Bødker, S., Constructing Common Information Spaces. Teoksessa ECSCW'97: Proceedings of the Fifth European Conference on Computer Supported Cooperative Work, Kluwer, 1997, 81-96.

[BaW97] Baldonado, M., Winograd, T., SenseMaker: An Information-Exploration Interface Supporting the Contextual Evolution of a User's Interests. Teoksessa Proceedings of the ACM Conference on Human Factors in Computing Systems, (CHI'97), ACM Press, 1997, 11-18.

[BeC92] Belkin, N., Croft, W., Information Filtering and Information Retrieval: Two Sides of the Same Coin? Communications of the ACM, 35, 12 (1992), 28-38.

[Ber97] Berners-Lee, T., Metadata Architecture. W3C, 1997.

<<http://www.w3.org/DesignIssues/Metadata.html>>

[BFM98] Berners-Lee, T., Fielding, R., Masinter, L., Uniform Resource Identifiers (URI): Generic Syntax. Internet Engineering Task Force, RFC 2396, 1998.

<<http://www.ietf.org/rfc/rfc2396.txt>>

[BGS01] Baudisch, P., Good, N., Stewart, P., Focus Plus Context Screens: Combining Display Technology with Visualization Techniques. Teoksessa Proceedings of the 14<sup>th</sup> Annual ACM Symposium on User Interface Software and Technology (UIST'01), ACM Press, 2001, 31-40.

[BjR00] Björk, S., Redström, J., Redefining the Focus and Context of Focus+Context Visualizations. Teoksessa IEEE Symposium on Information Visualization 2000 (INFOVIS'00), 9-10 October 2000, Salt Lake City, Utah, USA. IEEE Computer Society, 85-90.

[BMD99] Buckingham Shum, S., Motta, E., Domingue, J., Representing Scholarly Claims in Internet Digital Libraries: A Knowledge Modelling Approach. Teoksessa Abiteboul, S., Vercoustre, A-M., (toim.), Proceedings of ECDL'99: Third European Conference on Research and Advanced Technology for Digital Libraries, Paris, France, September 22-24, 1999. Springer-Verlag Lecture Notes in Computer Science, 423-442.

[BMD00] Buckingham Shum, S., Motta, E., Domingue, J., ScholOnto: An Ontology-Based Digital Library Server for Research Documents and Discourse. International Journal on Digital Libraries, 3, 3, (2000), 237-248.

[BrD96] Brown, J., Duguid, P., The Social Life of Documents. First Monday 1, 1, (1996).

<<http://www.firstmonday.dk/issues/issue1/documents/index.html>>

[BrG02] Brickley, D., Guha, R., RDF Vocabulary Description Language 1.0: RDF Schema. W3C Working Draft, 30 April, 2002.

<<http://www.w3.org/TR/2002/WD-rdf-schema-20020430>>

[Bud00] Budzik, J., et al., Beyond Similarity. Teoksessa Working notes of the AAAI 2000 Workshop on AI for Web Search, Austin, TX, USA, July 2000, AAAI Press.

[BuH94] Buckingham Shum, S., Hammond, N., Argumentation-based Design Rationale: What Use at What Cost? *International Journal of Human-Computer Studies*, 40, (1994), 603-652.

[BuH99] Budzik, J., Hammond, K., User Interactions with Everyday Applications as Context for Just-in-time Information Access. Teoksessa Proceedings of the 2000 International Conference on Intelligent User Interfaces, ACM Press, 2000, 44-51.

[Bus45] Bush, V., As We May Think. *Atlantic Monthly*, 176, 1, (1945), 101-108.  
<<http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>>

[BWD01] Boyack, K., Wylie, B., Davidson, G., Information Visualization, Human-Computer Interaction and Cognitive Psychology: Domain Visualizations. Sandia National Laboratories, VxInsight Project, 2001.  
<<http://vw.indiana.edu/visual01/boyack-et-al.pdf>>

[ByC01] Byun, H., Cheverst, K., Exploiting User Models and Context-Awareness to Support Personal Daily Activities. (Online) Proceedings of Workshop on User Modeling for Context-Aware Applications at the 8th International Conference on User Modeling (UM2001), Sonthofen, Germany, July 13, 2001.  
<<http://orgwis.gmd.de/~gross/um2001ws/papers/byun.pdf>>

[CaH00] Case, D., Higgins, G., How Can We Investigate Citation Behavior? A Study of Reasons for Citing Literature in Communication. *Journal of the American Society for Information Science*, 51, 7, (2000), 635-645.

[ChC99] Chen, C., Carr, L., Trailblazing the Literature of Hypertext: Author Co-Citation Analysis (1989-1998). Teoksessa HYPertext '99, Proceedings of the 10th ACM Conference on Hypertext and Hypermedia: Returning to Our Diverse Roots, ACM Press, 1999, 51-60.

[Che02] Chen, C., Information Visualization. *Information Visualization*, 1, 1, (2002), 1-4.

[ChP01] Chen, C., Paul, R., Visualizing a Knowledge Domain's Intellectual Structure. *IEEE Computer*, 34, 2, (2001), 65-71.

[CHR99] Carr, L., Hall, W., De Roure, D., The Evolution of Hypertext Link Services. *ACM Computing Surveys*, 31, 4es, (1999), Article No. 6.  
<<http://doi.acm.org/10.1145/345966.345992>>

[Coa01] Coates, T., et al., URIs, URLs, and URNs: Clarifications and Recommendations 1.0. Report from the Joint W3C/IETF URI Planning Interest Group. W3C, Note 21, September 2001.  
<<http://www.w3.org/TR/2001/NOTE-uri-clarification-20010921/>>

[CoB88] Conklin, J. Begeman, M., gIBIS: A Tool for Exploratory Policy Discussion. *ACM Transactions on Office Information Systems*, 6, 4, (1988), 303-331.

[Cob93] Cober, W., Contextual Constructivism: The Impact of Culture on the Learning and Teaching of Science. Teoksessa Tobin, K., (toim.), *The Practice of Constructivism in Science Education*. Hillsdale, NJ, Lawrence Erlbaum Associates, Inc., 1993, 51-69.

[Con87] Conklin, J., Hypertext: an Introduction and Survey. *IEEE Computer* 20, 9 (1987), 17-41.

[CPO01] Chen, C., Paul, R., O'Keefe, B., Fitting the Jigsaw of Citation: Information Visualization in Domain Analysis. *Journal of the American Society for Information Science and Technology*, 52, 4, (2001), 315-330.

[Cro00] Cronin, B., Bibliometrics and Beyond: Some Thoughts on Web-based Citation Analysis. *Journal of Information Science*, 27, 1, (2001), 1-7.

[DeH98] Dempsey, L., Heery, R., Metadata: A Current View of Practice and Issues. *Journal of Documentation*, 54 ,2 (1998), 145-172.

[Der96] Dervin, B., Given a Context by any Other Name: Methodological Tools for Taming the Unruly Beast. Teoksessa Vakkari, P., Savolainen, R., Dervin, B., (toim.), *Information Seeking in Context. Proceedings of an International Conference on Research in Information Needs, Seeking and Use in Different Contexts*, 14-16 August, 1996 Tampere, Finland, 13-38.

[DeW02] Dekkers, M., Weibel, S., Dublin Core Metadata Initiative Progress Report and Workplan for 2002. *D-Lib Magazine*, 8, 2, (2002).

<<http://www.dlib.org/dlib/february02/weibel/02weibel.html>>

[Dil01] Dillon, M., Metadata for Web Resources: How Metadata Works on the Web. Teoksessa *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium*, Library of Congress, 2001.

<[http://www.loc.gov/catdir/bibcontrol/dillon\\_paper.html](http://www.loc.gov/catdir/bibcontrol/dillon_paper.html)>

[Doy61] Doyle, L., Semantic Road Maps for Literature Searchers. *Journal of the ACM*, 8, 4, (1961), 553-578.

[Dub99] Dublin Core Metadata Element Set, Version 1.1: Reference Description. Dublin Core Metadata Initiative, DCMI Recommendation, 1999.

<<http://dublincore.org/documents/1999/07/02/dces/>>

[Dub00] Dublin Core Qualifiers. Dublin Core Metadata Initiative, DCMI Recommendation, 2000.

<<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>>

[Eng62] Engelbart, D., *Augmenting Human Intellect: a Conceptual Framework*. Summary Report, Stanford Research Institute, on Contract AF 49(638)-1024, October 1962.

<<http://www.histech.rwth-aachen.de/www/quellen/engelbart/ahi62index.html>>

[Fie99] Fielding, R., et al., Hypertext Transfer Protocol -- HTTP/1.1. Internet Engineering Task Force, RFC 2616, 1999.

<<http://www.ietf.org/rfc/rfc2616.txt>>

[Fis01] Fischer, G., User Modeling in Human-Computer Interaction. User Modeling and User-Adapted Interaction, 11, (2001), 65-86.

[Fon97] Foner, L., Yenta: A Multi-Agent, Referral Based Matchmaking System. Teoksessa Proceedings of the The First International Conference on Autonomous Agents (Agents'97), ACM Press, 1997, 301-307.

[FrC89] Frisse, M., Cousins, S., Information Retrieval from Hypertext: Update on the Dynamic Medical Handbook. Proceedings of ACM Hypertext '89, 1989, 199-211.

[Fre00] Freeman, L., Visualizing Social Networks. Journal of Social Structure, 1, 1, (2000).

<<http://zeeb.library.cmu.edu:7850/JoSS/article.html>>

[Gar55] Garfield, E., Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. Science, 122, 3159, (1955), 108-111.

[Gar79] Garfield, E., Is Citation Analysis a Legitimate Evaluation Tool? Scientometrics, 1, 4, (1979), 359-375.

[Gar96] Garfield, E., When to Cite. Library Quarterly, 66, 4, (1996), 449-458.

[GhG01] Ghidini, C., Giunchiglia, F., Local Models Semantics, or Contextual Reasoning = Locality + Compatibility. Artificial Intelligence, 127, (2001), 221-259.

[Gol92] Goldberg, D., et al., Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM 35, 12, (1992), 61-70.

[Gre94] Gresham, J., From Invisible College to Cyberspace College: Computer Conferencing and the Transformation of Informal Scholarly Communication

Networks. *Interpersonal Computing and Technology: An Electronic Journal for the 21st Century*, 2, 4, (1994), 37-52.

<<http://www.helsinki.fi/science/optek/1994/n4/gresham.txt>>

[Hel00] Dublin Core Metadata Element Set, Versio 1.1: Tallenteiden kuvailuformaatti. Helsingin yliopiston kirjasto, 2000.

<[http://www.lib.helsinki.fi/dublin\\_core/dcref-fin.html](http://www.lib.helsinki.fi/dublin_core/dcref-fin.html)>

[HeP96] Hearst, M., Pedersen, J., Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results. Teoksessa *Proceedings of the 19th Annual International Conference on Research and Development in Information Retrieval, ACM SIGIR'96*, 1996, 76-84.

[Her99] Hertzum, M., Six Roles of Documents in Professionals' Work. Teoksessa Bøfker, S., Kyng, M., Schmidt, K., (toim.), *Proceedings of the sixth European Conference on Computer-Supported Cooperative Work*, Copenhagen, Denmark, 1999, 41-60.

[Hil01] Hillman, D., Using Dublin Core. Dublin Core Metadata Initiative, DCMI Recommendation, 2001.

<<http://dublincore.org/documents/2001/04/12/usageguide/>>

[HiT85] Hiltz, S., Turoff, M., Structuring Computer-Mediated Communication Systems to Avoid Information Overload. *Communications of the ACM*, 28, 7, (1985), 680-689.

[HMM00] Herman, I., Melancon, G., Marshall, M., Graph Visualization and Navigation in Information Visualization: a Survey. *IEEE Transactions on Visualization and Computer Graphics*, 6, 1, (2000), 24-43.

[Hul01] Hull, D., *Science and Selection*. Cambridge University Press, 2001.

[Jos02] Joseph, S., NeuroGrid: Semantically Routing Queries in Peer-to-Peer Networks. Esitetty konferenssissa *International Workshop on Peer-to-Peer Computing*,

A Workshop Co-located with Networking 2002, Second International IFIP-TC6 Networking Conference, Pisa, Italy, May 19-24, 2002.

<<http://www.elet.polimi.it/p2p/papers/1.pdf>>

[Kob93] Kobsa, A., User Modeling: Recent Work, Prospects and Hazards. Teoksessa Schneider-Hufschmidt, M., Kühme, T., Malinowski, U., (toim.), Adaptive User Interfaces: Principles and Practice. North-Holland, Amsterdam, 1993, 111-128.

[Koh00] Kohonen, T., et al., Self Organization of a Massive Document Collection. IEEE Transactions on Neural Networks, 11, 3, (2000), 574-585.

[Kos98] Kostoff, R., The Use and Misuse of Citation Analysis in Research Evaluation. Scientometrics, 43, 1, (1998), 27-43.

[KoS02] Kokkelink, S., Schwänzl, R., Expressing Qualified Dublin Core in RDF/XML. Dublin Core Metadata Initiative, DCMI Proposed Recommendation, 2002.

<<http://dublincore.org/documents/2002/04/14/dcq-rdf-xml/>>

[KSM02] Kling, R., Spector, L., McKim, G., Locally Controlled Scholarly Publishing via the Internet: The Guild Model. The Journal of Electronic Publishing, 8, 1, (2002).

<<http://www.press.umich.edu/jep/08-01/kling.html>>

[KSS97] Kautz, H., Selman, B., Shah, M., Referral Web: Combining Social Networks and Collaborative Filtering. Communications of the ACM, 40, 3, (1997), 63-65.

[Lag02] Lagoze, C., et al., (toim.), The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 2.0 of 2002-06-14. Open Archives Initiative, 2002.

<<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>>

[LaN92] Lamming, M., Newman, W., Activity-Based Information Retrieval: Technology in Support of Personal Memory. Teoksessa Vogt, F., (toim.), Proceedings of Information Processing 92 (Personal Computers and Intelligent Systems), Vol. 3, Elsevier Science, Amsterdam, 1992, 68-81.



[LaV01] Lagoze, C., Van de Sompel, H., The Open Archives Initiative: Building a Low-barrier Interoperability Framework. Teoksessa Proceedings of the First ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL'01, ACM Press, 2001, 54-62.

[LBG99] Lawrence, S., Bollacker, K., Giles, C., Indexing and Retrieval of Scientific Literature. Teoksessa Eight International Conference on Information and Knowledge Management, CIKM'99, Kansas City, Missouri, November 2-6, 1999, 139-146.

[LeA00] Leuski, A., Allan, J., Details of Lighthouse. CIIR Technical Report IR-212, 2000.

<<http://ciir.cs.umass.edu/pubfiles/ir-212.pdf>>

[LGB99] Lawrence, S., Giles, C., Bollacker, K., Digital Libraries and Autonomous Citation Indexing. IEEE Computer, June 1999, 67-71.

[LFW01] Lieberman, H., Fry, C., Weitzman, L., Exploring the Web with Reconnaissance Agents. Communications of the ACM, 44, 8, (2001), 69-75.

[Lin97] Lin, X., Map Displays for Information Retrieval. Journal of the American Society for Information Science. 48, 1, (1997), 40-54.

[Low85] Lowe, D., Co-Operative Structuring of Information: The Representation of Reasoning and Debate. International Journal of Man-Machine Studies 23, 2, (1985), 97-111.

[Mae94] Maes, P., Agents That Reduce Work and Information Overload. Communications of the ACM, 37, 7, (1994), 31-40.

[MaM96] MacRoberts, M., MacRoberts, B., Problems of Citation Analysis. Scientometrics, 36, 3, (1996), 435-444.

[Mau96] Maurer, H., Hyper-G now Hyperwave, The Next Generation Web Solution. Addison-Wesley, 1996.

[McC87] McCarthy, J., Generality in Artificial Intelligence. *Communications of the ACM*, 30, 12, (1987), 1030-1035.

[Mer68] Merton, R., The Matthew Effect in Science. *Science*, 159, 3810, (1968), 56-63.

[MGT86] Malone, T., Grant, K., Turbak, F., The Information Lens: An Intelligent System for Information Sharing in Organizations. Teoksessa Mantei, M., Orbeton, P., (toim.), *Conference Proceedings on Human Factors in Computing Systems, (CHI'86)*, ACM Press, 1986, 1-8.

[Mil98] Miller, E., An Introduction to the Resource Description Framework. *D-Lib Magazine*, May 1998.

<<http://www.dlib.org/dlib/may98/miller/05miller.html>>

[Miz94] Mizruchi, M., Social Network Analysis: Recent Achievements and Current Controversies. *Acta Sociologica*, 37, (1994), 329-343.

[MRC95] Mackinlay, J., Rao, R., Card, S., An Organic User Interface for Searching Citation Links. Teoksessa Katz, I., et al. (toim.), *Conference Proceedings on Human Factors in Computing Systems, (CHI'95)*, ACM Press/Addison-Wesley Publishing Co., NY, USA, 1995, 67-73.

[Muk99] Mukherjea, S., Information Visualization for Hypermedia Systems. *ACM Computing Surveys*, 31, 4es, (1999), Article No. 6.

<<http://doi.acm.org/10.1145/345966.345984>>

[Nej02] Nejd, W., et al., EDUTELLA: A P2P Networking Infrastructure Based on RDF. Teoksessa *Proceedings of The Eleventh International Conference on World Wide Web, WWW2002*, ACM Press, 2002, 604-615.

<<http://www2002.org/CDROM/refereed/597/index.html>>

[Nel99] Nelson, T., The Unfinished Revolution and Xanadu. *ACM Computing Surveys*, 31, 4es, (1999), Article No. 37.

<<http://doi.acm.org/10.1145/345966.346039>>

[NeM92] Newman, S., Marshall, C., Pushing Toulmin Too Far: Learning From an Argument Representation Scheme. Xerox PARC Technical Report No. SSL-92-45, 1992.

<<http://www.csd.tamu.edu/~marshall/toulmin.pdf>>

[OaK98] Oard, D., Kim., J., Implicit Feedback for Recommender Systems. Teoksessa Kautz, H., (toim.), Papers from the 1998 AAAI Workshop on Recommender Systems, Technical Report WS-98-08, The AAAI Press, 1998, 80-82.

[Obi01] Obitko, M., Ontologies - Description and Applications. Czech Technical University in Prague, Gerstner Laboratory for Intelligent Decision Making and Control Series of Research Reports, Report No: GL 126/01. 2001.

<<http://cyber.felk.cvut.cz/gerstner/reports/GL126.pdf>>

[Obj02] The Common Object Request Broker: Architecture and Specification, version 3.0. Object Management Group, 2002.

<<http://www.omg.org/cgi-bin/doc?formal/02-06-33>>

[Pag98] Page, L., et al., The PageRank Citation Ranking: Bringing Order to the Web. Technical report, Stanford, Santa Barbara, CA 93106, January, 1998.

<<http://www-db.stanford.edu/~backrub/pageranksub.ps>>

[PaM02] Pant, G., Menczer, F., Myspiders: Evolve Your Own Intelligent Web Crawlers. Autonomous Agents and Multi-Agent Systems, 5, 2, (2002), 221-229.

[Pea89] Pearl, A., Sun's Link Service: A Protocol for Open Linking. Teoksessa Proceedings of ACM Hypertext'89. 1989, 137-146.

[Pla98] Plassard, M-F., (toim.), Functional Requirements for Bibliographic Records, Final Report. International Federation of Library Associations (IFLA) UBCIM Publications - New Series, Vol. 19. 1998.

<<http://www.ifla.org/VII/s13/frbr/frbr.pdf>>

[Pop72] Popper, K., Objective Knowledge – An Evolutionary Approach. Oxford University Press, 1972.

[ReV97] Resnick, P., Varian, H., Recommender Systems. Communications of the ACM, 40, 3, (1997), 56-58.

[RhM00] Rhodes, B., Maes, P., Just-in-time Information Retrieval Agents. IBM Systems Journal, 39, 3&4, (2000), 685-704.

[SAW94] Schilit, B., Adams, N., Want, R., Context-aware Computing Applications. Teoksessa Proceedings of the Workshop on Mobile Computing Systems and Applications, IEEE, 1994, 85-90.

[Sch92] Schatz, B., Building an Electronic Community System. Journal of Management Information Systems, 8, 3, (1991-92), 87-107.

[Sel99] Selvin, A., Supporting Collaborative Analysis and Design with Hypertext Functionality. Journal of Digital information, 1, 4, (1999).

<<http://jodi.ecs.soton.ac.uk/Articles/v01/i04/Selvin/>>

[ShM99] Shipman, F., Marshall, C., Formality Considered Harmful: Experiences, Emerging Themes, and directions on the Use of formal representations in Interactive Systems. Computer-Supported Cooperative Work, 8, 4, (1999), 333-352.

[SKR99] Schafer, J., Konstan, J., Riedl, J., Recommender Systems in E-Commerce. Teoksessa Proceedings of the ACM Conference on Electronic Commerce (EC-99), 1999, 158-166.

[Sma73] Small, H., Co-citation in the Scientific Literature: A New Measure of the Relationship Between Two Documents. Journal of the Americal Society for Information Science, July-August, 1973, 265-269.

[Sma99] Small, H., Visualizing Science by Citation Mapping. Journal of the American Society for Information Science, 50, 9, (1999), 799-813.

[SmG85] Small, H., Garfield, E., The Geography of Science: Disciplinary and National Mappings. *Journal of Information Science*, 11, (1985), 147-159.

[Swa64] Swanson, D., Dialogues with a Catalog. *Library Quarterly* 34, (1964), 113-125.

[SYV01] Singh, M., Yu, B., Venkatraman, M., Community-based Service Location. *Communications of the ACM*, 44, 4, (2001), 49-54.

[Tri83] Trigg, R., A Network-Based Approach to Text Handling for the Online Scientific Community. Ph.D. dissertation, University of Maryland Technical Report, TR-1346, 1983.

[TwN98] Twidale, M., Nichols, D., A Survey of CSCW for Digital Libraries. Technical Report CSEG/4/98, Computing Department, Lancaster University, 1998.  
<<http://www.comp.lancs.ac.uk/computing/research/cseg/projects/ariadne/docs/survey.html>>

[WHK97] Wahl, M., Howes, T., Kille, S., Lightweight Directory Access Protocol (v3). Internet Engineering Task Force, RFC 2251, 1997.  
<<http://www.ietf.org/rfc/rfc2251.txt>>

[WhW97] White, M., Wang, P., A Qualitative Study of Citing Behavior: Contributions, Criteria and Metalevel Documentation Concerns. *Library Quarterly*, 67, 2, (1997), 122-154.

[YVS99] Yu, B., Venkatraman, M., Singh, M., A Multiagent Referral System for Expertise Location. Teoksessa Working Notes of the AAAI Workshop on Intelligent Information Systems, Orlando, FL, 1999, 66-69.