

Hypermedia ja visualisointi

Pietari Koskela

Helsinki 22.2.2002

Hypermediajärjestelmät –seminaari

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

Hypermedia ja visualisointi

Pietari Koskela

Hypermediajärjestelmät –seminaari

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

18.2.2002, 12 sivua

Kun www sivut alkoivat yleistyä ensimmäiset ihmiset surffailivat verkossa ja lukivat kaikkea mitä vastaan tuli. Nopeasti kuitenkin tietomäärä kasvoi niin suureksi ettei kaikkea ollut enää mielekästä lukea. Tiedon etsimisestä tuli erittäin hankalaa vaikka tieto itsessään oli jossain verkon syövereissä. Yhä useammin hukkuminen tiedon valtatiellä synnytti tarpeen hakukoneille. Sivustojen kasvaessa myös sivustojen sisäinen tieto pitää järjestää. Tähän tarkoitukseen on kehitetty erilaisia sivukarttoja. Hypermedia ja sen visualisointi sisältävät paljon tutkimatonta ja avoimia kysymyksiä joihin kehitettyjä ratkaisuja tullaan käsittelemään.

Avainsanat: hyperteksti, hypertekstin visualisointi, www sivukartta

Sisältö

Sisältö

1	Johdanto	1
2	WWW sivukartat.....	3
3	Sivukartta menetelmiä.....	6
4	Yhteenveto	9
	Lähteet.....	10

1 Johdanto

Hypermedia, hypertexti, multimedia. Näiden raja on häilyvä mutta voitaneen sanoa, että multimedia ei ole interaktiivista. Voi olla, että multimedia yhtenä interaktiivisena osana isompaa kokonaisuutta voi olla hypermediaa. Hypermedia on useimmille tuttu hypertextin muodossa, jotka arkipäiväisessä elämässä konkretisoituu www-sivujen muodossa. Hypermedian tunnuspiirteet täytyvät kun käyttäjät vuorovaikutteisesti luovivat www sivustoilla.

Klikkailemalla hypertextiin sisällytetyjä linkkejä pääsee uuteen hypertextiin joka voi sisältää linkin taas johonkin uuteen hypertextiin ja niin edelleen. Hypertextin vahvuus kaikessa laajuudessaan ja joustavuudessaan on myös sen heikkous. Kaikki tietävät, että Internetillä ei ole alkua eikä loppua. Www sivustolla voi olla alku ja loppu, mutta sen ollessa osa suurempaa kokonaisuutta, hypertextin alku ja loppu hämärtyy. Hypertextin luonteesta johtuen lukijat kokevat kokonaisuuden hallinnan vaikeaksi. Lukiessa jostain aiheesta on vaikea sanoa, onko jo lukenut kaiken aiheeseen liittyvän vai ei. Lukijan seurattessa tekstissä olevia linkkejä, ei aina tiedä onko samassa artikkelissa kuin aikaisemmin vai onko siirtynyt johonkin toiseen www sivustoon. Hyvin laaditun www sivuston sivukarttaa (site map) katsellessa lukija voi saada selvän käsityksen sivuston laajuudesta, tämä sivusto voi kuitenkin olla vain hypertextin osa, ja kokonaiskuvaa on näin ollen mahdotonta saada. Kirjasta näkee heti, ennen kuin on kerinnyt edes koskea siihen, kuinka paksu kirja on. Kirjan paksuudesta voi hieman päätellä kuinka paljon siinä on tekstiä, toisin kuin hypertextin kyseessä ollessa.

Mukherjea [Muk99] kiteyttää hyvin osan hypermedian ongelmista: ”Huomattava ongelma hypermediajärjestelmissä on eksyä hyperavaruuteen, koska ei ole olemassa rajoituksia siitä, kuinka käyttäjä navigoi informaatioavaruudessa.”

Tietomäärän kasvaessa lukijan on yhä vaikeampi löytää etsimäänsä hypertextin joukosta.

Hakukoneet ovat osoittautuneet erittäin hyödyllisiksi etsittäessä tietoa koko Internetistä.

Sivuston rakennetta ja tiedon etsintää sivustoilla on huomattu parhaiten helpottavan sivukartat.

Koska hypertexti on hypermedian yleisin muoto, on sivukartat varsin keskeisessä osassa tässä seminaarityössä.

Hypermediaksi voitaisiin myös kuvitella kasvojentunnistus. Ensin määriteltäisiin kasvojen parametrit esimerkiksi Jian-Kangin [Wu97] esittämän algoritmin mukaan. Tämän jälkeen saatuja kasvoja voitaisiin muuttaa tekemällä valintoja kahden eri kasvonpiirteiden mukaan valitsemalla lähempänä haluttua kasvonpiirteitä oleva kuva.

Interaktiiviset elokuvat ovat jääneet kokeiluasteelle, tällaiset elokuvat joissa katsojan valinnat vaikuttavat elokuvan kulkuun ja lopputulokseen, voitaisiin pitää myös hypermediana.

Luvussa 2 käsitellään sivukarttoja yleisesti ja niiden vaatimuksia ja odotuksia. Luvussa 3 on esitelty muutamia sivukarttojen menetelmiä.

2 WWW sivukartat

Kuten jo luvussa 1 todettiin, ovat sivukartat osoittautuneet hyödyllisiksi lukijan kokonaisuuden hahmottamisen helpottamisessa. David ja Paul ovat koonneet artikkelissaan MAPA: järjestelmä hierarkian visualisointiin ja verkkosivustoille lisäämiseen [DuK98], sivukartan kymmenen tärkeintä ominaisuutta.

1. Näyttää sivuston rakenteen ja tarjoaa yleiskuvan sivuston tärkeimmistä osista. yleiskuvan luominen vaatii karsimaan osan kaikista mahdollisista saavutettavista osista.
2. Näyttää missä käyttäjä on. Sivukartalta tulee käydä ilmi missä kohtaa sivustoa käyttäjä on.
3. Näyttää missä käyttäjä on jo käynyt. Sivukartalta pitää nähdä missä hän on jo käynyt ja missä on vielä käymättä. Käytyjen sivujen ei saa rajoittua vain yhteen käyntikertaan, vaan myös aikaisemmillä kerroilla käytyjen sivujen tulee olla käytyjen sivujen joukossa.
4. Näyttää mihin tästä voi mennä. Sivukartalla pitää näkyä joukko linkkejä joita tällä hetkellä olevalta sivulta voidaan seurata.
5. Näytä ryhmien välinen yhteys. Sivukartalla pitää näkyä sivuston samankaltaisten sivujen yhteydet toisiinsa.
6. Näytä kuinka tänne on jouduttu. Sivukartalla pitää näkyä kuinka sivuston sisällä on liikuttu ja päästy nykyiselle sivulle.
7. Näytä kuinka monta käyttäjää on käynyt tällä sivulla. Eteenkin ylläpitäjiä, mutta myös käyttäjiä kiinnostaa kuinka moni on käynyt tällä sivulla ja kuinka moni on sivulla juuri nyt.
8. Sivukartan tulisi erotella kiinnostavat sivut. Sivukartan tulisi näyttää käyttäjälle käyttäjää kiinnostavat sivut. Kuten on helppo ymmärtää tämän ominaisuuden toteuttaminen on erittäin vaikeata toteuttaa.
9. Sivukartan pitää näyttää paljon tietoa hyvin pienessä tilassa. Sivukartta joka vaatii paljon tilaa ei ole käytännöllinen. Sivukartan automaattiseen luomiseen ei saa kulua liian paljon aikaa.
10. Sivukartan pitää toimia verkkoselaimessa ilman lisäohjelmistoa tai pitkiä latausaikoja. Eteenkin satunnaisien käyttäjien kannalta on erittäin tärkeää, että sivukartta on nopeasti käytettävissä. Usein sivustoa käyttäville ja ylläpidolle kertalataus ei ole merkityksellinen, koska se tehdään vain kerran.

Yllä oleva lista on sängen kattava esitys sivukartan vaatimuksista. Ajan kuluessa ja uusien menetelmien syntyessä listaan voitaisiin varmasti lisätä useita kohtia, ja osa listassa olevista kohdista osoittautuu tarpeettomiksi, silti listan sisältämät kohdat tuntuvat tällä hetkellä varsin uskottavilta.

Riippumatta itse sivukartan luomiseen käytettävästä menetelmästä, näyttäisi kaikkien menetelmien noudattavan ainakin osittain tiettyä kaavaa. Isojen tietomäärien visualisointi on vaikeaa ja liian isojen tietomäärien visualisointi käy mahdottomaksi. Yleistämällä visualisoitavaa tietomäärää joukko saadaan pienemmäksi ja visualisointi on mahdollista. Yleistämiseen ja yksinkertaistamiseen on olemassa useita menetelmiä [MuF95], joita eri sivukartta menetelmät hyödyntävät.

Suodattaminen (filtering) on yksi keino jolla hypermedian tietojoukkoa voidaan pienentää. Suodattaminen on käytännöllistä kun visualisoinnin tulos on liian suuri tai liian pieni. Jos visualisoinnin tulos on liian suuri voidaan tulosjoukkoa suodattaa lisää tekemällä tarkentavia suodatuksia saadusta tulosjoukosta. Liian pienen tulosjoukon seurauksena voidaan hakuetoja suodattaa ja näin saada aikaan laajempi tulosjoukko.

Abstraktio on keino hypermediatietojoukon pienentämiseksi. Tarkoituksena on kasata samantyyppiset asiat yhteen korkeamman tason asioiksi. Esimerkiksi: päärynä, banaani ja mandariini yksinkertaistetaan hedelmäksi. abstraktio on johonkin pisteeseen asti hyvin kuvaava ja voi jopa tuoda lisää tietoa suppeammassa tietojoukossa. ylitettyään kriittisen pisteen abstraktio hukkaa tietoa erittäin nopeasti ja näin ollen tulosjoukko käy käyttökelvottomaksi. Abstraktion kaltainen kategorisointi on tietoa lajitteleva menetelmä. Peterin, Jamesin ja Raolin artikkelissa [PPR96] kerrotaan kuinka www sivuja voidaan lajitella erilaisten arvojen perusteella. Esimerkiksi sivun koon mukaan , eli kuinka monta bittiä sivu vie. Sisään ja ulos menevien linkkien määrää voidaan pitää myös yhtenä kategorisoinnin mittana. Lasketaan kuinka monta linkkiä sivulle johtaa tai montako linkkiä sivulta on muualle. Latausten määrä on hakukoneiden käytössä oleva kategorian mittari, sillä lasketaan kuinka monta kertaa sivu on ladattu kyseisellä ajanjaksolla. Sivun syvyys sivustossa kategorisoi sivun sen mukaan kuinka monta /-merkkiä sivun osoitteessa on.

Aikaisemmin listassa kohta kahdeksan, jossa ohjelma arvaa korostaa käyttäjää kiinnostavat linkit tuntuu olevan yleinen tutkimisen aihe ja ongelma. Tätä aihetta tutkitaan voimakkaasti myös hakukoneiden näkökannalta katsottuna. Hakukoneet voidaan ajatella eräinä Internetin globaaleina sivukarttoina, jotka tuottavat käyttäjän antamien suodattimien kautta kuvan tietoverkosta. Koska suodatettu tietojoukko on usein vielä erittäin laaja on Lycos:n [PiP97] tapaisia hakukoneita kehitetty arvaamaan käyttäjää kiinnostavat sivustot ja näyttämään ne tulosjoukon alussa.

Erityyppisiä visualisointi tekniikoita voidaan kategorioida , kuten David ja Paul ovat artikkelissaan [DuK98] tehneet.

- Graaffipohjainen (graph based) visualisointi raakatekstistä ryhmitellään linkit jokin järkevä mallin mukaan. Mallissa käy ilmi selkeät keskinäiset suhteet.
- Verkon rakenne (web structures) visualisointi raakatekstistä on yleisin verkon omainen esitys siitä, millainen sivuston hierarkia tiedostojärjestelmässä on. Tämä malli kasvaa helposti liian isoksi, että se olisi helposti ymmärrettävissä. Tällainen malli ei aina tue käyttäjän näkemää sivustoa ja voi sekoittaa käyttäjää jopa lisää.
- Hierarkkinen rakenne on rajoitettu olemassa oleva graafi. Hyvänä puolena on, että hierarkialla on selvä kanta isä- lapsisuhteista. Hierarkkinen rakenne on hyvä luokiteltaessa kokonaisuutta ja jäykkyydestään huolimatta auttaa käyttäjää ymmärtämään sitä helpommin.
- Asyklinen rakenne (acyclic structure) luodaan hierarkkisesta mallista. Asyklinen malli sallii moniperinnän (multiple inheritance) ja on käytännöllinen varsinkin pienillä joukoilla. Rakenne tulee kuitenkin erittäin monimutkaiseksi ja vaikeasti ymmärrettäväksi joukon ollessa suuri.
- Tila rakenne (Spatial structures) käyttää tiedettyjä koordinaatti järjestelmiä navigointipisteiden esittämiseen. Rakenteen akselit valitaan jokin merkittävä akseli, jolle navigointipisteet sijoitetaan.
- Naapurisuus (Neighborhood) ovat tila rakenteen alitapaus, koska siltä puuttuu oma koordinaatisto. Vapaaseen koordinaatistoon sijoitetaan joukon osat niiden sisältämien yhteisten olioiden mukaan.

3 Sivukartta menetelmiä

Sivukartta menetelmät voidaan jakaa myös edellistä lukua karkeammin kahteen eri ryhmään, menetelmiin jotka tuottavat graafisen lopputuloksen ja menetelmiin joiden lopputulos ei ole graafinen.

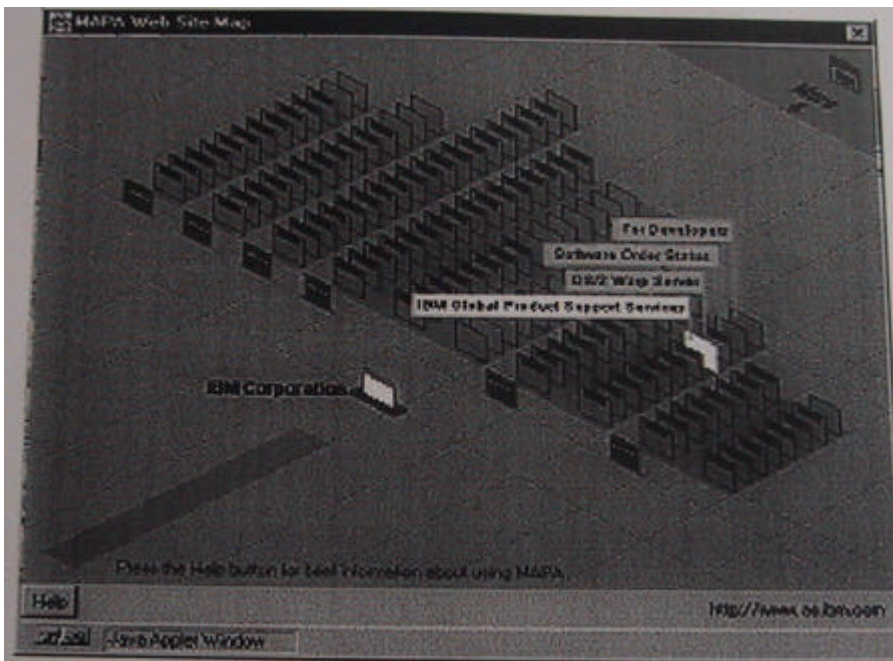
Ei graafisten sivukarttojen toteutusta pidetään yleisesti helpompana ja sivukartan luomisessa käytettävät menetelmät ovat pohjimmiltaan yksinkertaisempia. Ei graafisten sivukarttojen etuna on usein, että niitä katsellessa ei tarvitse ladata mitään erillisiä sivukartan näyttämishjelmiä. MBnetin sivukartta [MBnWWW] sisältää myös grafiikkaa muutaman kuvan muodossa, mutta lukeutuu kuitenkin hyvänä esimerkkinä ei graafisten sivukarttojen sarjaan [Kuva1.]. Sivukartta on tyyppillinen ja hyvin yleinen. Tällaisen sivukartan ylläpito manuaalisesti ja ensikertaa luominen ovat vaivattomia verrattuna moniin muihin sivukartta menetelmiin.



Kuva 1, Mbnetin sivukartta. <http://www.mbnet.fi/sivukartta/>

George [Fur97] kuvaa hyvin listamaisia sivukarttoja jollaista Mbnetin sivukartta kuvassa 1 edustaa. Lisäämällä hieman grafiikkaa ja toiminnallisuutta sivukartta saadaan täyttämään huomattavasti enemmän sivukartan vaatimuksia. Halutunlainen symboli voidaan asettaa sivukarttaan siihen kohtaan jossa käyttäjä tällä hetkellä on. Piirtämällä nuolia käyttäjän nykyisestä kohdastaan sivukartalla, voidaan ilmaista mihin käyttäjän on mahdollista navigoida sivulta. Listaan voidaan piirtää myös käyttäjän historiapolku. Samaan ryhmään kuuluvia sivuja voidaan korostaa joko ympäröimällä ne laatikoin tai käyttäen pohjaväriä sivun nimen alla. Tämän tyyppisissä sivukartoissa on vaarana, että niiden suhteellinen paikka sivukartassa ei anna oikeaa kuvaa sivun luonteesta. Sivukartan osien kuvilla harvoin on mitään todellista tarkoitusta .

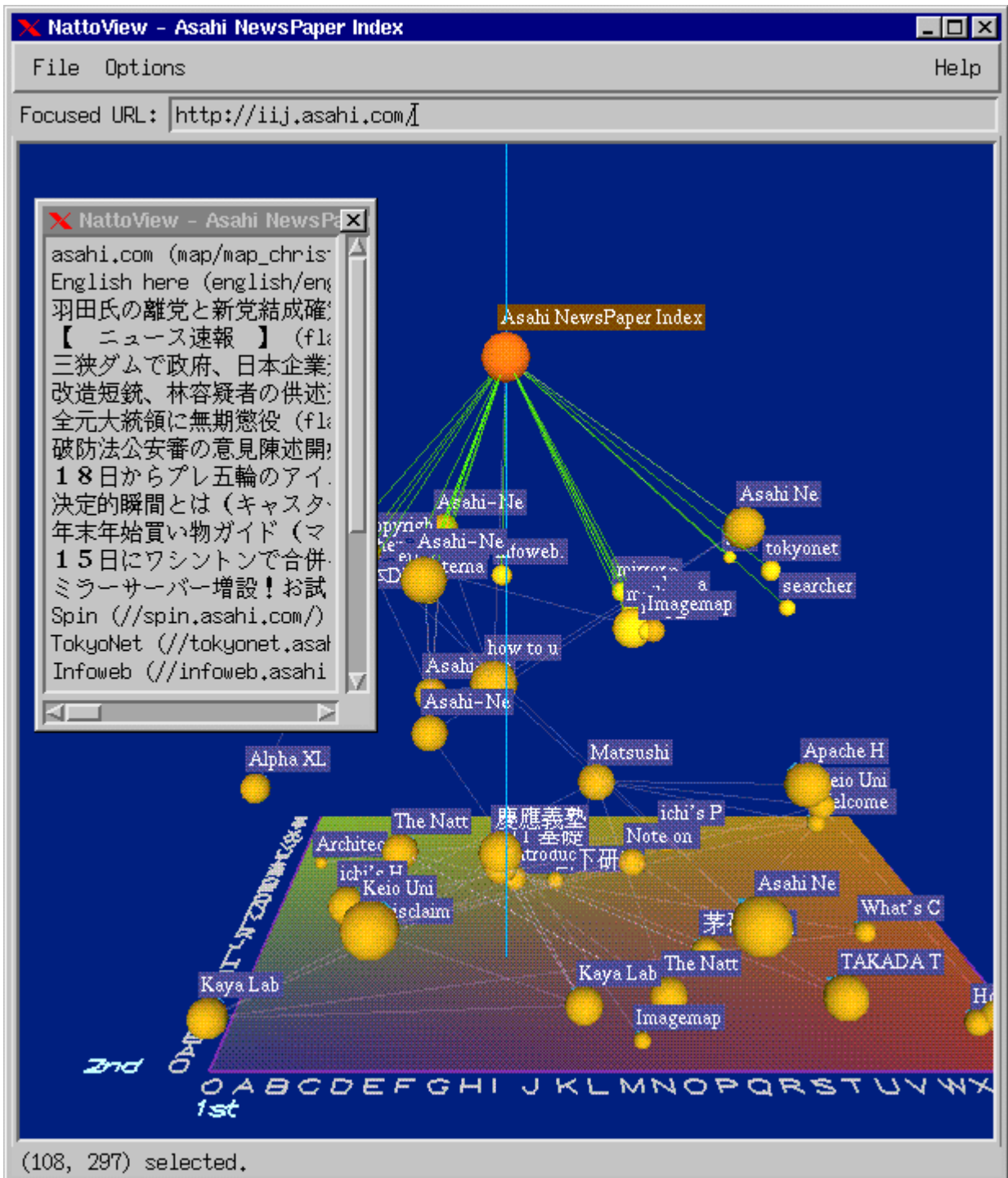
Graafisia sivukarttamenetelmiä löytyy huomattavasti enemmän kuin tekstipohjaisia. Graafiset sivukartat voitaneen jakaa kahteen ryhmään, 2d sivukartat ja 3d sivukartat. 3d ulottuvuuden ajatuksena on että käyttäjä lentää sivukarttamallin lävitse ja ymmärtää näin sen paremmin. MAPA [DuK98],[Ben99],[] on yksi sivukarttamenetelmä joka tuottaa sivukartan jossa on mukana useita 3d elementtejä. MAPA:a kehitettäessä ideoita otettiin muun muassa perinteisistä kartoista. Kuvassa 2 on MAPA:lla tehty sivukartta. Kuvassa eri sivut on kuvattu laatikolla jota klikkaamalla saadaan lisää informaatiota halutusta sivusta. Kuvakulma on yläviistosta, kuten varsinkin vanhemmissa kaupunkikartoissa oli tapana. Jokainen sivu on omassa kategoriassaan, jotka on eroteltu toisistaan erilaisilla pohjaväreillä.



Kuva 2, MAPA sivukartta.

MAPA luo sivukartan 500-50 000 sivun sivustosta. MAPA:n rakentamalla sivukartan käyttäminen on helpompaa kuin 3d käännösten teko 3d avaruudessa.

Yksi 3d avaruuden tekevä menetelmä on Natto [Ben00] ,[NatWWW]. Kuten kuvan 3 perusteella voidaan kuvitella, että suuressa sivustossa kokonaiskuva ei ehkä ole niin selkeää kuin voisi olla. Navigointi 3d avaruudessa vaatii varmasti aluksi sxpaljon totuttelua.



Kuva 3, Natto sivukartta <http://www.myo.inst.keio.ac.jp/groups/IPS/NattoView/>

4 Yhteenveto

Sivukartoilla on merkittävä osuus hypertekstin visualisoinnissa, se nopeuttaa ja helpottaa lukijaa ymmärtämään luettavan tekstin paikkaa suuremmissa kokonaisuuksissa ja antaa kuvan missä mennään. Sivukarttojen kehittyessä ja niiden toimintaperiaatteen yleistyessä sivukartat voivat nykyistä paremmin täyttää niille soveltuvat tehtävät. Selainten yhteensopivuus ja lisäosien (plugin) parantuessa, saavat myös sivukarttojen toteuttajat paremmat työvälineet parempien sivukarttojen toteuttamiseen.

Lähteet

- [Ben00] Benford, S. ja kumppanit, Three dimensional visualization of the world wide web. ACM computing surveys symposium on hypertext, 2000.
- [DuK98] Durand, D., ja Kahn, P., MAPA(tm):a system for inducing and visualizing hierarchy in websites, ACM Hypertext 98, Pittsburgh PA USA,66-76
- [dynWWW] Mapping websites, <http://www.dynamicdiagrams.net/>
- [Fur97] Furnas, G. W., effectice view navigation, ACM CHI 97, Atlanta GA USA, maaliskuussa 1997
- [MBnWWW] Mbnetin sivukartta. <http://www.mbnet.fi/sivukartta/>
- [MuF95] Mukherjea, S. ja Foley, S., Visualizing the world-wide web with the navigational view builder, Computer networks and ISDN systems. Special issue on the third international world-wide web conference, Darmstadt Germany, huhtikuussa 1995, 1075-1087
- [Muk99] Mukherjea, S., Information Visualization for Hypermedia Systems. ACM Computing Surveys 31,4 (1999).
- [NatWWW] The Natto view,<http://www.myo.inst.keio.ac.jp/groups/IPS/NattoView/>
- [PiP97] Pitkow, J., Pirolli, P., Life,death and lawfullness on the electronic frontier , ACM CHI 97, Atlanta GA USA, maaliskuussa 1997, 383-390
- [PPR96] Pirolli, P., Pitkow, J., Rao, R., Silk from a sow's ear: Extracting usable structures from the web, ACM SIGCHI 96, Vancouver, Canada, toukokuussa 1996, 118-125

- [Wu97] Wu, J.-K., *Transactions on knowledge and data engineering*. Content-Based indexing of Multimedia Databases, *IEEE*, vol. 9 no. 6 november/december 1997, 978-989