

hyväksymispäivä

arvosana

arvostelija

## **Hypermedian esittäminen yleisnäkömön kautta**

Teemu Sidoroff

Helsinki 29.10.2004

Seminaaritutkielma

HELSINGIN YLIOPISTO

Tietojenkäsittelytieteen laitos

## Hypermedian esittäminen yleisnäkömön kautta

Tässä seminaaritutkielmassa käsitellään hypertekstin esittämistä yleisnäkömön kautta esittelemällä järjestelmiä, jotka helpottavat erilaisten suurten aineistojen hahmottamista ja tiedon etsintää. Aluksi tarkastellaan näkömöhakua, joka mahdollistaa suuren aineiston selaamisen usean eri dimension kannalta. Relation Browser++ on näkömöhakua hyödyntävä järjestelmä, joka tarjoaa käyttöliittymän isojen homogeenisten aineistojen selaamiseen. WebTOC-järjestelmä generoi sivustolle navigointipalkin, joka näyttää sivuston aihealueiden lisäksi myös sivuilla olevien mediatyyppien jakautuman. Lopuksi esitellään Site Browser, joka on tarkoitettu web-sivustojen selaamisen koko WWW:n mittakaavassa.

ACM Computing Classification System (CCS):  
H.4.0 [Information Systems applications]: General

# Sisältö

<b>1 Johdanto</b>	<b>1</b>
<b>2 Yksittäisten dokumenttikokoelmien hallinta</b>	<b>1</b>
2.1 Näkymäselaus . . . . .	2
2.2 Relation Browser++ . . . . .	4
2.3 WebTOC . . . . .	5
<b>3 Site Browser</b>	<b>7</b>
3.1 Site Browserin tarjoama näkymä web-sivustoista . . . . .	7
3.2 Järjestelmän toiminta-alusta . . . . .	8
<b>4 Yhteenveto</b>	<b>9</b>
<b>Lähteet</b>	<b>10</b>

# 1 Johdanto

Tiedon löytäminen suurista tietomassoista on haastavaa. WWW:ssä hakukoneet ovat tehokas apuväline tiedonhakuun, mutta pelkkään sanahakuun perustuvat palvelut eivät kaikkiin tiedonhakutehtäviin sovellu. Käyttäjä ei esimerkiksi aina osaa ilmaista tiedontarvettaan selkeästi, jolloin hakulausekkeiden pukeminen sanoiksi on hankalaa. Hakukoneet eivät myöskään tarjoa käyttäjälle selkeää yleiskuvaa haun kohteena olevasta aineistosta, vaan ainoastaan yksittäisiä aineiston tietoalkioita, toisin sanoen yksittäisiä HTML-sivuja tai muita dokumentteja. Tiedon selaaminen on hyvä vaihtoehto suoralle tiedonhauille kun halutaan saada yleiskuva aiheesta tai aineistosta [Mar95]. Selaamisella, joka on ollut WWW:n ja muiden hypertekstijärjestelmien perustana alunperinkin [BL99], on siis edelleen käyttötarkoituksensa tiedonhaussa.

Suurien aineistojen kohdalla on kuitenkin haastavaa tarjota kattavaa esitystä koko aineiston sisällöstä. Koska suuren aineiston selaaminen ei ole tehokasta käyttäjän nopean väsymisen ja huomion herpaantumisen helppouden takia [Mar95], on aineistosta tiivistettävä käyttäjälle yleiskuva, joka kiteyttää aineiston tärkeimmät aspektit. Tämän yleiskuvan avulla käyttäjä voi muodostaa aineistosta käsityksen nopeasti ja vaivattomasti.

Tässä seminaaritutkielmassa käsitellään järjestelmiä ja menetelmiä, jotka tarjoavat mahdollisuuden tarkastella suuria web-aineistoja yleisnäkömön kautta. Luvussa 2 tarkastellaan yksittäisten dokumenttikokoelmien, yleensä yksittäisen web-sivuston, hallintaan kehitettyjä ratkaisuja ja esimerkkijärjestelmiä. Luvussa 3 tarkastellaan mittakaavaltaan astetta suurempia aineistoja varten kehitetty Site Browser, jonka tavoitteena on toimia koko WWW:n kattavana yleisnäkömäläimena.

## 2 Yksittäisten dokumenttikokoelmien hallinta

Tässä osassa keskitytään homogeenisten hypermedia-kokoelmien esittämisessä käytettyihin tekniikoihin ja järjestelmiin. Aluksi osassa 2.1 esitellään näkömöhaku (*faceted browsing*), jota on hyödynnetty monilla web-sivustoilla ja jossain määrin myös muissa tässä tutkielmassa käsiteltävissä järjestelmissä. Osissa 2.2 ja 2.3 käsitellään kahta Java-sovelmana toimivaa järjestelmää. Osassa 2.2 esiteltävä Relation Browser++ tarjoaa selaus- ja hakukäyttöliittymän suurille tietokannoista tuotetuille homogeenisille aineistoille ja osan 2.3 WebTOC-järjestelmä puolestaan muodostaa web-

sivuston rakenteen pohjalta sivustolle eräänlaisen sisällysluettelon.

## 2.1 Näkymäselaus

Suuret aineistot on etsimisen ja selauksen mahdollistamiseksi yleensä luokiteltu jonkin dimension mukaan. Esimerkiksi kirjastoissa kirjat jaotellaan hierarkisen Deweyn desimaalijärjestelmän perusteella aihealueittain. Näkymäselauksessa aineisto on luokiteltu yhden dimension sijaan useamman ortogonaalisten, toisistaan irrallisten, dimensioiden mukaan. Usean dimension luokittelu mahdollistaa aineiston selaamisen eri näkökulmien perusteella ja antaa käyttäjällä useampia mahdollisuuksia löytää yksittäinen tietoalkio [H<sup>+</sup>02].

Näkymäselaus muodossa tai toisessa on käytössä useilla eri web-sivustoilla, erityisesti verkkokaupoissa. Yksi esimerkki tällaisesta näkymäselauksista hyödyntävästä kaupasta on IBM:n kannettavia mikroja myyvä verkkokauppa<sup>1</sup>, joka tarjoaa käyttäjälle mahdollisuuden etsiä mielestään mikroa valitsemalla sille käyttäjän haluamia ominaisuuksia vapaavalintaisessa järjestyksessä. Eri ominaisuuksia, kuten muistin määrää, hintaa, painoa ym. valitsemalla hakutulosten määrä tarjolla olevista mikroista karsiutuu. Näkymäselauksen ansiosta käyttäjä voi vapaavalintaisessa järjestyksessä valita tärkeäksi näkemäänsä ominaisuuksia kunnes hakutulosten riittävän alhaisella tasolla, jotta tuloksia on mielekästä selata.

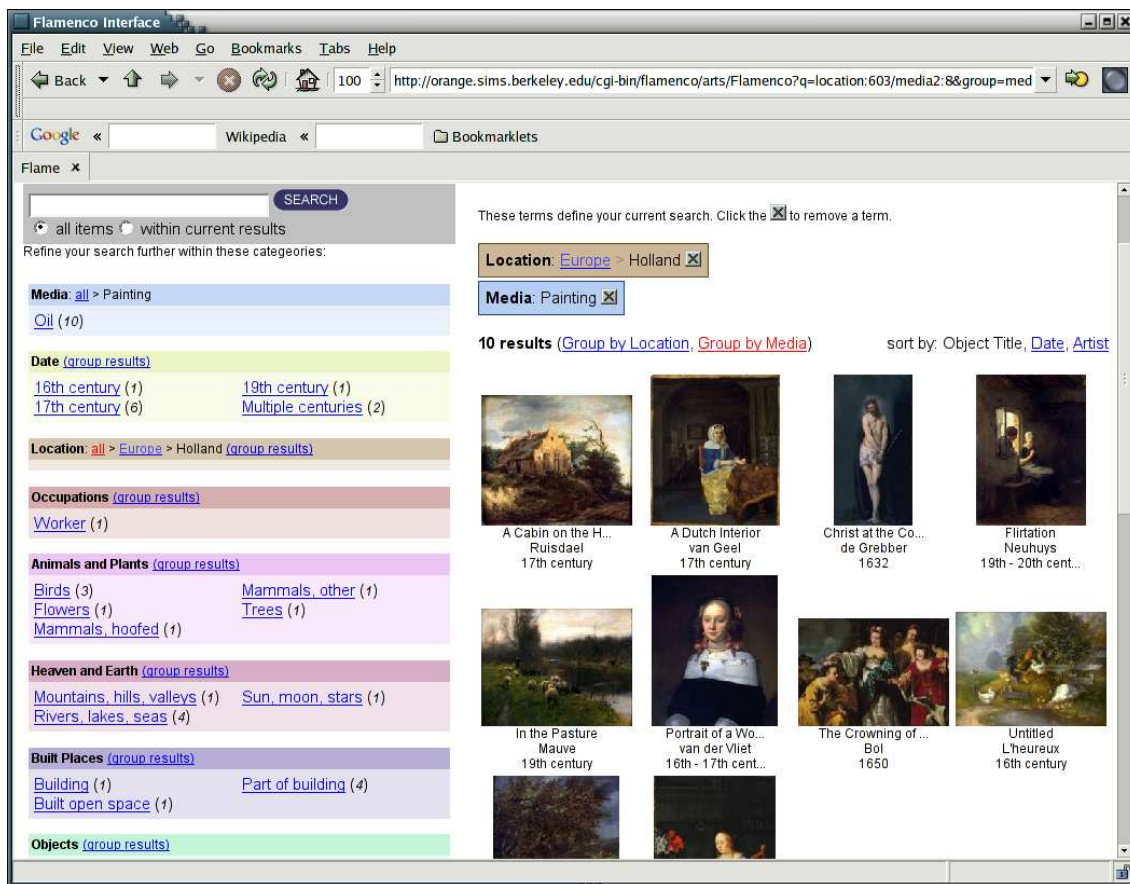
Berkeley'n yliopiston Flamenco-projekti [FLA] on soveltanut näkymähakua ja kehittänyt siihen perustuvaa järjestelmää tiedonhaun apuvälineeksi suuria aineistoja sisältäville web-sivustoille. Järjestelmää on kehitettäessä lähtökohtana on ollut erityisesti käytettävyyden [H<sup>+</sup>02]. Seuraavassa tarkastellaan projektin esittämää ratkaisumallia aineiston hallintajärjestelmäksi.

Flamenco-projektin Lee et al. vertaavat suureen tietomassaan tutustumista kolmevaiheiseen shakkipeliin, johon kuuluu avausvaihe, pelin keskivaihe ja pelin loppuvaihe [Y<sup>+</sup>03]. Erotuksena shakkipeliin analogiassa on se, että vaiheita ei tarvitse käydä läpi järjestyksessä vaan niiden välillä voi siirtyä vapaasti vaiheesta toiseen.

Avausvaiheessa järjestelmä tarjoaa käyttäjälle yleisnäkyvän aineistosta, jossa on esitettyinä kaikkien dimensioiden ylimmän tason kategoriat, joiden mukaan aineisto on jaoteltu. Tässä vaiheessa käyttäjä saa yleiskuvan järjestelmän tietosisällöstä ja aihealueista. Valitessaan yhden esitetyistä kategorioista lähempään tarkasteluun siirtyy käyttäjä pelin keskivaiheeseen.

---

<sup>1</sup><http://www.ibm.com/products/us/>



Kuva 1: Flamencon käyttöliittymä selauksen keskivaiheessa. Ruudun vasemmalla puolella näkyy aineiston kategorisointi ja oikealla puolella hakutulokset. Hakutulosten yläpuolella näkyy tuloksen tuottaneet valitut kategoriat.

Keskivaiheessa käyttäjälle esitetään ruudun oikeassaa laidassa hakutulos, eli kaikki valittuun kategoriaan kuuluvat tietoalkiot (ks. kuva 1). Ruudun vasemmassa laidassa ovat näkyvillä kaikki ne kategoriat, jotka liittämällä nykyiseen hakulausekkeeseen eivät tuota tyhjää hakutulosta. Käyttäjä voi valita hakulausekkeeseen uusia esillä olevia kategorioita. Uusia kategorioita liittämällä hakulausekkeeseen tuloksena palautetaan kategorioiden leikkauksen sisältämät tietoalkiot, eli ne tietoalkiot jotka kuuluvat kaikkiin hakulausekkeessa mukana oleviin kategorioihin. Lisäksi kategorioita voi poistaa hakulausekkeesta ja näin laajentaa tulosta.

Kun käyttäjä valitsee tarkasteltavaksi yhden tietoalkion siirrytään käyttöliittymässä analogiassa esitettyyn loppupeliin. Loppupelissä esitetään yksittäiseen tietoalkioon (tietoalkio voi olla esim. taideteos tai valokuva) liittyvät metatiedot.

Näkymäselaus tarjoaa käyttäjälle useita eri polkuja etsittävään tietoon ja se sovel-

tuu suurien, yhtenäisten aineistojen hallintaan. Näkymäselausjärjestelmän toteuttaminen vaatii aineistolta kuitenkin paljon metadataa. Lisäksi aineiston on oltava tarpeeksi homogeenistä, jotta kaikki tietoalkiot saadaan luokiteltua samojen luokarakenteiden mukaan.

## 2.2 Relation Browser++

Relation Browser++ (RB++) [Z<sup>+</sup>04, MB03] on järjestelmä, joka tarjoaa selaus- ja hakukäyttöliittymän suurelle datajoukkoille. RB++ tarjoaa näkymän suoraan tietokannassa olevaan datajoukkoon, eikä se tässä mielessä ole suoraan ole sidottu hypertextiaineistoihin. Järjestelmään tutustuminen on kuitenkin perusteltua, sillä suuri osa hypertextinä esitetystä tiedosta tuotetaan tietokannoista ja RB++-järjestelmää voidaan käyttää hyväksi erityisesti hypertextinä esitettyyn aineistoon tutustumisessa.

The screenshot shows the 'UNC Movie Collection' Java Applet Window. It features three filter menus: Genre, Decade, and Format. The Genre menu has 'Film-Noir' selected with a count of 68. The Decade menu has '1940-1949' selected with a count of 68. The Format menu has 'VHS' selected with a count of 65. Below the filters, there are buttons for 'Restart', 'Fewer Categories <<', and 'More Categories >>'. The main area displays a table with 68 results. The table has columns for Title, Date, and Director. A tooltip is visible over the 'Date' column, showing the selected filters: Genre: Film-Noir,Thriller, Decade: 1940-1949, Format: DVD,Laser Disc,VHS. The table lists movies such as 'Letter' (1940, William Wyler), 'Shanghai Gesture' (1941, Josef von Sternberg), 'High Sierra' (1941, Raoul Walsh), 'Maltese Falcon' (1941, John Huston), 'I Wake Up Screaming' (1941, Humberstone, H. Bruce), 'This Gun For Hire' (1942, Frank Tuttle), 'Shadow of a Doubt' (1942, Alfred Hitchcock), 'Glass Key' (1942, Stuart Heisler), 'Double Indemnity' (1944, Billy Wilder), 'Laura' (1944, Otto Preminger), 'Woman in the Window' (1944, Fritz Lang), 'Murder, My Sweet [AKA: Farewell My Lovely]' (1944, Edward Dmytryk), 'Phantom Lady' (1944, Robert Siodmak), 'Mildred Pierce' (1945, Michael Curtiz), 'House on 92nd Street' (1945, Henry Hathaway), 'Ministry of Fear' (1945, Fritz Lang), 'Scarlet Street' (1945, Fritz Lang), 'Conflict' (1945, Curtis Bernhardt), 'Postman Always Rings Twice' (1946, Tay Garnett), 'Stranger' (1946, Orson Welles), 'Notorious' (1946, Alfred Hitchcock), 'Strange Love of Martha Ivers' (1946, Lewis Milestone), 'Big Sleep (Hawks/Bogart)' (1946, Howard Hawks), 'Gilda' (1946, Charles Vidor), and 'Spiral Staircase' (1946, Robert Siodmak). At the bottom, a red text string reads '(Genre=Film-Noir) AND (Decade=1940-1949)\*\*\*\*'.

Title	Date	Director
Letter	1940	Wyler, William
Shanghai Gesture	1941	von Sternberg, Josef
High Sierra	1941	Walsh, Raoul
Maltese Falcon	1941	Huston, John
I Wake Up Screaming	1941	Humberstone, H. Bruce
This Gun For Hire	1942	Tuttle, Frank (I)
Shadow of a Doubt	1942	Hitchcock, Alfred
Glass Key	1942	Heisler, Stuart
Double Indemnity	1944	Wilder, Billy
Laura	1944	Preminger, Otto
Woman in the Window	1944	Lang, Fritz
Murder, My Sweet [AKA: Farewell My Lovely]	1944	Dmytryk, Edward
Phantom Lady	1944	Siodmak, Robert
Mildred Pierce	1945	Curtiz, Michael
House on 92nd Street	1945	Hathaway, Henry
Ministry of Fear	1945	Lang, Fritz
Scarlet Street	1945	Lang, Fritz
Conflict	1945	Bernhardt, Curtis
Postman Always Rings Twice	1946	Garnett, Tay
Stranger	1946	Welles, Orson
Notorious	1946	Hitchcock, Alfred
Strange Love of Martha Ivers	1946	Milestone, Lewis
Big Sleep (Hawks/Bogart)	1946	Hawks, Howard
Gilda	1946	Vidor, Charles
Spiral Staircase	1946	Siodmak, Robert

Kuva 2: Relation Browser++ käyttöliittymän esittämä näkymä elokuva-arkistoon. Jokainen tietue toimii linkkinä elokuvasta kertovalle web-sivulle.

RB++-järjestelmä on toteutettu Java-sovelmana ja sen käyttöliittymä käyttää hyväksi näkymäselauksesta tuttuja ideoita. Järjestelmä tarjoaa käyttäjälle mahdollisuuden rajata selattavaa datajoukkoa eri dimensioihin kuuluvien attribuuttien perusteella. Datajoukosta voidaan valita kiinnostavia tietoalkioita, joita voidaan siirtyä tarkastelemaan lähemmin tietoalkiota tietoalkioita vastaavilta web-sivuilta.

Esimerkiksi datajoukko, joka koostuu elokuvista voidaan järjestelmässä jaotella elokuvien formaatin, lajityypin ja tuotantovuoden mukaan. Käyttäjä voi rajata datajoukosta osajoukon valitsemalla järjestelmän avulla vaikkapa ne elokuvat, jotka kuuluvat lajityypin *film noir* ja jotka on valmistettu vuosina 1940-1949 (ks. kuva 2). Valittuihin elokuviin voi tutustua lähemmin valitsemalla niistä yhden tuloslistalta, jolloin käyttäjä pääsee tutkimaan kyseisen elokuvan tietoja tietokannassa olevasta tietueesta generoidulta web-sivulta.

RB++-järjestelmän käyttöliittymässä on pyritty helppokäyttöisyyteen. Järjestelmä visualisoi automaattisesti hiirtä attribuutin päälle siirrettäessä kuinka suuri datajoukko käyttäjän valitsemilla attribuuteilla tuotetaan. Järjestelmä tarjoaakin hyödyllisen työkalun tietokannoista generoitujen web-sivujen selaamiseen ja erityisesti kiinnostavien tietoalkioiden osajoukkojen etsimiseen sellaisilta sivustoilta, jotka eivät tarjoa tehokkaita hakutoimintoja.

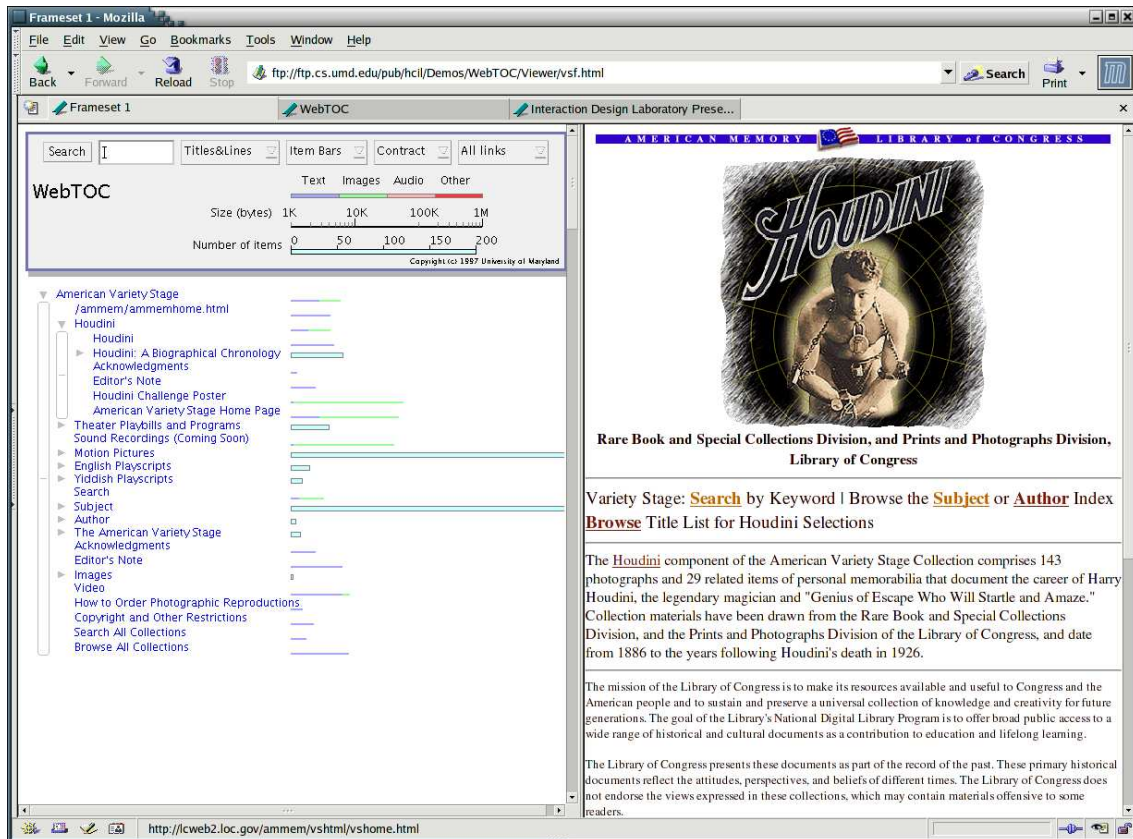
## 2.3 WebTOC

WebTOC [N<sup>+</sup>97] on järjestelmä, jolla voidaan generoida web-sivustoista sisällysluetteloita ja joita voidaan käyttää eräänlaisina sivuston navigointipalkkina. Järjestelmä koostuu kahdesta komponentista, parserista joka generoi sisällysluettelon ja Java-sovelmasta joka visualisoi tämän sisällysluettelon.

WebTOC generoi sisällysluettelon käymällä läpi sivuston sisäiset linkit ja generoi malla jokaisesta linkistä uuden hierarkiatason. Linkkien päässä olevasta informaatiosta tallennetaan nimen lisäksi myös linkin takana olevan solmun rakenne ja koko kilotavuina. Rakenteella tarkoitetaan solmussa esiintyvien mediatyyppien (teksti, kuvat, ääni, ja muut data) suhdetta.

Generoitua sisällysluetteloä hyödynnetään visualisoimalla se WebTOC-järjestelmään kuuluvan Java-sovelman avulla. Sovelma avaa sisällysluettelon selattavan sivuston rinnalle erilliseen kehykseen. Sisällysluettelo on esitetty puumaisena rakenteena, jonka laajennettavia solmuja hyödyntämällä käyttäjä voi nopeasti tutustua sivuston sisältöön (ks. kuva 3). Jokaisen solmun yhteydessä on esitetty myös solmun koos-





Kuva 3: WebTOC-järjestelmällä augmentoitu web-sivusto.

tumus, eli siinä olevien kuvien ja muiden multimediaelementtien suhde tekstiin, horisontaalisena palkkina. Sisällysluettelon solmut toimivat hyperlinkkeinä sivuille, joten sisällysluetteloä voi käyttää sisällön selaamisen lisäksi myös sivustolla navigoimiseen.

WebTOC-järjestelmän käyttökelpoisuudessa on ilmeisiä puutteita. Järjestelmä ei generoi sisällysluetteloä automaattisesti, joten esitetty sisällysluettelo voi vanhentua helposti. Lisäksi käyttäjän on rakennettava sisällysluettelo jokaisesta sivustosta erikseen. WebTOC-järjestelmässä esitetyt ideat ovat kuitenkin kehityskelpoisia. Vaikka usealla sivustolla on navigointipalkki, WebTOC-järjestelmän generoima navigointimahdollisuus sivustolla auttaa käyttäjää löytämään sisältöä, joka ei sivun alkuperäisestä asusta välttämättä ole helposti löydettävissä. Lisäksi sisällysluettelossa esitetyt graafit solmujen tietomäärästä ja mediatyypijakaumista ovat informatiivisia.

Olisikin mielenkiintoista nähdä WebTOC-järjestelmän kaltainen selainlaajennokse-na toteuttu työkalupalkki joka automaattisesti generoisi selatun sivuston rakenteesta sisällysluettelon selaimen sivupalkkiin. Tällaisia selainlaajennoksia on mahdollista

tehdä esimerkiksi Mozilla-selaimiin.

## 3 Site Browser

Edellä esitetyt järjestelmät rajoittuvat ainoastaan yhden sivuston yleisnäkymän tarkasteluun. Site Browser [Gib04] on järjestelmä, joka tarjoaa yleisnäkymän useista sivustoista. Sen tavoitteena on antaa käyttäjälle kuva web-sivustojen rakenteesta ja sisällöstä yhtenäisen käyttöliittymän kautta, sekä mahdollistaa WWW:n selaus sivustotasolla sivutason sijaan.

Tässä luvussa tarkastellaan ensin osassa 3.1 Site Browserin toiminta-ajatusta ja osassa 3.2 järjestelmän toiminta-alustaa.

### 3.1 Site Browserin tarjoama näkymä web-sivustoista

David Gibson käyttää osassa 2.1 käsitellyn Flamenco-projektin esittelemää shakkipeli-metaforaa esitellessään Site Browserin toiminta-ajatusta [Gib04]. Muiden yleisnäkyviä tarjoavien järjestelmien tapaan Site Browser keskittyy pelin aloitusvaiheeseen, jossa aineistoon saadaan ensikosketus, sekä keskivaiheeseen, jossa pyritään hahmottamaan selattavasta aineistosta informaatiota jostain tietystä näkökulmasta.

Site Browserin avulla käyttäjä pystyy selaamaan WWW:tä sivustotasolla tavanomaisen sivutason sijaan. Site Browserin tapauksessa sivustolla tarkoitetaan kaikkia yksittäisellä palvelimella sijaitsevia sivuja. Järjestelmä kerää sivustoista tilastotietoa, joka esitetään käyttäjälle erilaisina näkyminä. Näkymien avulla käyttäjä pystyy tarkastelemaan sivuston eri attribuutteja. Näitä attribuutteja ovat:

- sivuston koko. Koko ilmaistaan sivujen lukumääränä, sivuilla olevien sanojen lukumääränä sekä sivuston tiedostojen tavumääränä.
- Sivujen attribuutit. Sivusta tallennetaan niiden kielikoodi, HTTP:n palauttama koodi ja tieto sivuilla olevista linkeistä multimediatiedostoihin (ääni- ja videotiedostoihin).
- HTTP-palvelimen palauttamien tietojen päiväyksistä.
- Sivuston sivujen URL:ien muodostama hakemistorakenne.

- Muille sivustoille osoittavat linkit ja näiden lukumäärät. Linkeistä tallennetaan ainoastaan sivuston osoite, ei tarkkaa URL:ia.
- Sivuilla esiintyvien sanojen frekvenssit.

Nämä attribuutit esitetään käyttöliittymän neljässä eri näkymässä. Sivuston koko ja sivujen attribuutit esitetään yleisnäkymässä. Hakemistorakenne esitetään selattavana puurakenteena, samaan tapaan kuin esimerkiksi Microsoftin Windows Explorer tiedostohallintajärjestelmässä. Sivuston avainsanoista esitetään niiden esiintymissivujen lukumäärän mukaan järjestetty lista. Avainsanoista selviää nopeasti sivuston aihealue.

Linkit toisille sivustoille esitetään listana, joka on järjestetty sivustolle osoittavien linkkien esiintymismäärän mukaan. Sivuston yleisimmät linkit paljastavat paljon sivuston luonteesta. Esimerkiksi akateemisilla sivustoilla linkit johtavat usein toisille akateemisille sivustoille. Sivuston linkit johtavat usein myös maantieteellisesti sivuston lähialueella sijaitseville sivustoille, joten niiden perusteella voidaan muodostaa veikkaus myös sivuston maantieteellisestä sijainnista.

Linkkinäkymässä esitettyjä linkkejä voidaan seurata, jolloin Site Browser esittää yleisnäkymän sivustolta, jolle seurattu linkki osoitti. Tämän näkymän kautta Site Browserilla voidaan siis selata WWW:tä sivustotasolla, jolloin järjestelmän eri näkymien avulla saadaan nopeasti käsitys jokaisen selatun sivuston tarkoituksesta, rakenteesta, toteutuksesta ja aihealueesta.

## 3.2 Järjestelmän toiminta-alusta

Edellisessä osassa käsitellyt Site Browserin sivustoista esittämät attribuutit vaativat raskasta laskentaa. Site Browser on rakennettu IBM:n Almadenin tutkimuskeskusten Web Fountain -klusterin [D<sup>+</sup>03, Gib04] päälle. Web Fountain tarjoaa skaalautuvan alustan erilaisille suurien aineistojen tekstianalyysiohjelmistoille.

Site Browser hyödyntää Web Fountainin tarjoamaa verkkoindeksoijakomponenttia (*web crawler*) joka kykenee varastoimaan järjestelmän 100 teratavua säilytystilaa sisältävään tietokantaan 40 miljoonaa sivua päivässä. Indeksoijan keräämät sivut analysoidaan. Analysoinnissa sivusta varastoidaan edellisessä kappaleessa esitetyt attribuutit. Sivuston sivujen tiedot yhdistetään sulautusvaiheessa sivuston yhteenvedoksi joka tallennetaan Site Browserin käyttöön.

Web Fountainin massiivisen tallennuskapasiteetin ansiosta Site Browserin avulla pystytään selaamaan sivustoja reaaliaikaisesti. Lähes kaikki sivustot joita järjestelmän koekäyttäjät selasivat, olivat indeksoitu Web Fountainiin ja näin suoraan Site Browserin saavutettavissa.

Site Browser ei ole siis aivan kaikkien käyttäjien ulottuvilla, vaan siinä käytetyt WWW:n analysointimenetelmät vaativat tehokkaan suoritusalustan. Tällainen järjestelmä voisi kuitenkin toimia keskitettynä web-pohjaisena palveluna, jolloin se tarjoaisi uuden mielenkiintoisen tason web-selaukseen.

## 4 Yhteenveto

Tässä seminaaritutkielmassa tarkasteltiin erityisesti WWW:ssä sijaitsevien suurten tietovarastojen tarpeita varten kehitettyjä yleisnäköselauksen menetelmiä ja järjestelmiä. Havainnollinen yleisnäkö nä auttaa suuren tietomassan hahmottamisessa ja toimii hyvänä lähtökohtana tiedohaussa.

Näköselaukseen on hyödynnetty useissa yleisnäköjärjestelmissä sekä web-sivustojen hakuominaisuuksissa. Näköselauksen avulla sivuston tietoalkioita voidaan selata verrattaen vapaavalintaisessa järjestyksessä, eikä se sido käyttäjää noudattamaan yksittäistä hakupolkua tiedon löytämiseen. Näköselauksen ideoita on hyödynnetty myös Relation Browser++-järjestelmässä, joka voidaan liittää tietokannasta generoitavien web-sivustoihin selaukseen liittämiseksi.

WebTOC on järjestelmä, joka muodostaa sivustolle sisällysluettelon sen rakenteen ja sisällön perusteella. WebTOC-järjestelmän avulla saadaan nopeasti yleiskäsitys sivuston sisällöstä. Järjestelmä on hyvä apuväline varsinkin sellaisia sivustoja selatessa, joilta puuttuu selkeä sivukartta tai navigaatiopalkki.

Site Browser on edellisiä kunnianhimoisempi järjestelmä. Sen avulla käyttäjä pystyy selaamaan WWW:tä sivustotasolla tavanomaisen sivutason sijaan. Site Browserin alustana toimii tehokas tekstianalyysiin kehitetty klusteri, joka sisältää suuren tietovaraston, johon Site Browser voi tallentaa tuhansien sivustojen yhteenvetoja.

Kaikissa edellä esitetyissä järjestelmissä ongelmana on se, että ne vaativat joko järjestelmän erillistä soveltamista tai asentamista yksittäiselle sivustolle tai Site Browserin tapauksessa aineiston esikäsittelyä ja suurta laskenta- ja varastointikapasiteettia.

## Lähteet

- BL99 Berners-Lee, T., *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. Harper San Francisco, 1999.
- D<sup>+</sup>03 Dill, S. et al., SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation. *Proceedings of the Twelfth International Conference on World Wide Web*, Budapest, Unkari, toukokuu 2003.
- FLA Flamenco Project. <http://bailando.sims.berkeley.edu/flamenco.html>. [21.10.2004]
- Gib04 Gibson, D., The Site Browser: Catalyzing Improvements in Hypertext Organisation. *Proceedings of the fifteenth ACM conference on Hypertext & hypermedia*. ACM Press, aug 2004.
- H<sup>+</sup>02 Hearst, M. et al., Finding the Flow in Web Site Search. *Communications of The ACM*, 45,9(2002).
- Mar95 Marchionini, G., *Information seeking in electronic environments*. Cambridge University Press, 1995.
- MB03 Marchionini, G. ja Brunk, B., Toward a General Relation Browser: A GUI for Information Architects. *Journal of Digital Information*, 4,1(2003).
- N<sup>+</sup>97 Nation, D. A. et al., Visualizing Web Sites using a Hierarchical Table of Contents Browser: WebToc. *Proceedings of the third conference on Human Factors and the Web*, kesäkuu 1997.
- Y<sup>+</sup>03 Yee, K.-P. et al., Faceted Metadata for Image Search and Browsing. *Proceedings of the conference on Human factors in computing systems*. ACM Press, huhtikuu 2003.
- Z<sup>+</sup>04 Zhang, J. et al., Relation Browser++: an information exploration and searching tool. *Proceedings of the The National Conference on Digital Government Research*, toukokuu 2004.