

hyväksymispäivä arvosana

arvostelija

Hyperwave-järjestelmä

Matti Pihlaja

Helsinki 3.12.2004

Seminaaritutkielma

HELSINGIN YLIOPISTO
Tietojenkäsittelytieteen laitos

Tiedekunta/Osasto — Fakultet/Sektion — Faculty		Laitos — Institution — Department	
Matemaattis-luonnontieteellinen		Tietojenkäsittelytieteen laitos	
Tekijä — Författare — Author			
Matti Pihlaja			
Työn nimi — Arbetets titel — Title			
Hyperwave-järjestelmä			
Oppiaine — Läroämne — Subject			
Tietojenkäsittelytiede			
Työn laji — Arbetets art — Level		Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Seminaaritutkielma		3.12.2004	12 sivua + 0 liitesivua
Tiivistelmä — Referat — Abstract			
<p>Seminaaritutkielma antaa yleiskuvan Hyperwave-järjestelmästä (ent. Hyper-G). Hyperwave on www-yhteensopiva tietokantapohjainen Internet-palvelin, joka on varustettu www:hen nähden uusilla ominaisuuksilla. Hyperwave on suunniteltu erityisesti hypertekstimuotoisen tiedon hallinnointiin. Arkkitehtuurin ohella tutkielmassa käsitellään lyhyesti Hyperwaven historiaa ja nykytilaa.</p> <p>ACM Computing Classification System (CCS): H.5.4 [Information Interfaces and Presentation]: Hypertext/Hypermedia</p>			
Avainsanat — Nyckelord — Keywords			
Hypermedia, Hyper-G, Hyperwave			
Säilytyspaikka — Förvaringsställe — Where deposited			
Muita tietoja — Övriga uppgifter — Additional information			

Sisältö

1 Johdanto	1
2 Hyperwaven historiaa	2
3 Dokumenttimalli ja linkitys	3
4 Hyperwave-palvelin	5
4.1 Asiakas-palvelin -malli	5
4.2 Palvelimen arkkitehtuuri	7
4.3 Tekstihaku aineistosta	9
5 Hyperwave nykyään	9
6 Yhteenveto	10
Lähteet	12

1 Johdanto

Hyperwave [Mau96] www-palvelin on hierarkisella tavalla dokumentteja tallentava palvelinjärjestelmä. Tekijät kutsuvat Hyperwavea toisen sukupolven ratkaisuksi, www:n ollessa ensimmäistä sukupolvea. Hyperwave on yhteensopiva ww:n kanssa. Tämä tarkoittaa sitä, että olemassa olevissa selaimilla, kuten Mozilla, voidaan selata Hyperwave-palvelimen dokumentteja. Hyperwave on tehty laajentamaan www:tä tuomalla siihen uusia ominaisuuksia. Näillä ominaisuuksilla pyritään ratkaisemaan www:ssä havaittuja erinäisiä puutteita.

Www-ympäristössä ei pystytä valvomaan linkkirakenteen eheyttä, jonka johdosta dokumenttien linkit vanhenevat sitä mukaa, kun dokumentteja siirretään paikasta toiseen. Hyperwavessa kaikki objektit tallennetaan palvelinkohtaiseen tietokantaan. Dokumentit on erotettu linkkirakenteesta siten, että tiedot linkeistä tallennetaan omaan rakenteeseensa. Koska kaikki muutokset linkitykseen tehdään Hyperwave-palvelimen kautta, Hyperwave voi siten valvoa linkkirakennetta.

Käyttötutkimusten mukaan www-käyttäjälle on usein epäselvää, missä suhteessa nykyinen dokumentti on muihin nähden. Tätä ongelmaa Hyperwavessa koitetaan ratkaista dokumenttien hierarkisella rakenteella ja erilaisilla hypermediaa visualisovilla malleilla. Lisäksi www:stä puuttuu dokumenttien versionhallinta. Käyttäjä näkee kyllä linkistä, onko hän käynyt linkin viittaamassa dokumentissa (tämä voidaan hoitaa esim. värityksen avulla). Linkin väritys ei kuitenkaan muutu, jos dokumentin sisältö muuttuu. Hyperwavessa dokumenttien muutosaika kirjataan ylös ja sitä voidaan käyttää hyväksi linkin statuksen määrittämisessä.

Suurena puutteena tekijät mainitsevat sen, että www:ssä ei ole sisäänrakennettua tukea rajatulle tiedonhauulle, vaan tiedonhaku suoritetaan ulkopuolisten tarjoajien hakukoneiden avulla. Tällöin hakukoneen kohteena on koko verkko. Nykyään tosin esim. Googella [Goo04] haku voidaan rajata koskemaan tiettyä pal-

velinta. Hyperwave-järjestelmässä tekstihakuja voidaan rajata koskemaan tiettyä Hyperwave-palvelinta tai dokumenttikokoelmaa. Maurer on esittänyt myös muita ongelmia www:hen liittyen, mutta kaikkien näkökohtien käsittely muodostaisi tämän tutkielman kannalta liian kattavan kokonaisuuden.

Hyperwave-järjestelmä kuvataan tässä pääosin sen mukaan, miten se on esitetty Hermann Maurerin kirjassa *Hyperwave - The Next Generation Web Solution* [Mau96]. Luvussa 2 käsitellään Hyperwaven historiaa. Hyperwave dokumenttimallia ja dokumenttien välistä linkitystä käsitellään luvussa 3 ja palvelimen arkkitehtuuria luvussa 4. Lopuksi tehdään vielä katsaus Hyperwaven nykytilaan luvussa 5.

2 Hyperwaven historiaa

Hyperwave (ent. Hyper-G) kehitettiin alunperin Itävallassa Grazin yliopistossa professori Hermann Maurerin johdolla [Mau96]. Hyperwave-projekti sai rahoituksen Itävallan Tieteen ministeriöltä vuonna 1990 ja Frank Kappe teki tohtorinväitöskirjassaan suunnitelman Hyperwaven arkkitehtuurista vuonna 1991. Jo Hyperwaven ensimmäinen versio kykeni keskustelemaan www-palvelimien kanssa, sillä Hyperwaven ohjelmoivat olivat yhteydessä www:n suunnittelijoihin CERN:ssä. Hyperwaven ja www:n kehittyminen tapahtuikin melko samanaikaisesti.

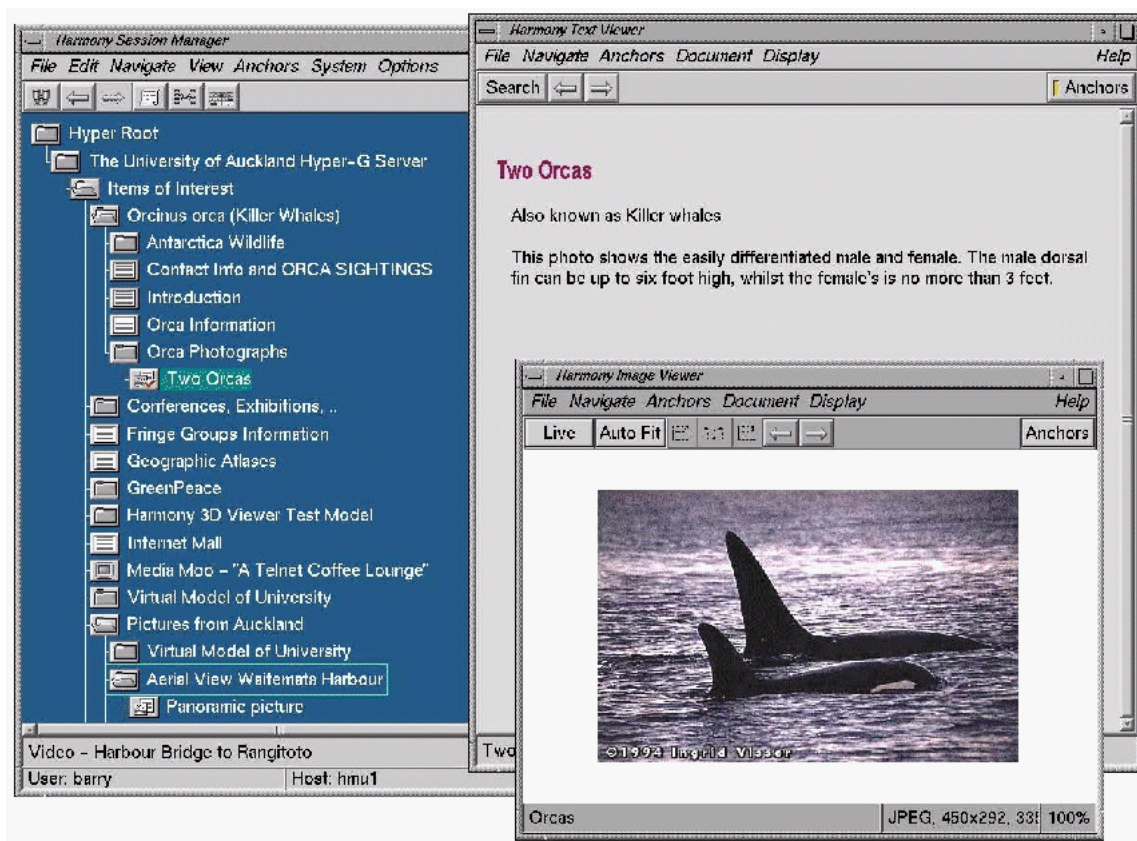
Positiivisten käyttökokemusten jälkeen Grazin Yliopistossa Hyperwavesta päätettiin tuottaa kaupallinen versio. Vuonna 1992 Euroopan Avaruusjärjestö ESA otti "Guide and Directory-järjestelmänsä perustaksi Hyperwaven. Windows-pohjainen Hyperwave-selain Amadeus julkaistiin vuotta myöhemmin ja tätä seurasi Unix-alustalla toimiva Harmony. Vastaavasti NCSA:ssa (National Center for Supercomputing Applications) kehitettiin Mosaic-selain Unixille 1993 alkupuolella.

3 Dokumenttimalli ja linkitys

Hyperwaven dokumenttirakenteen ja linkkien kuvaaminen perustuu Hyperwave-kirjan lukuihin 9 ja 11 [Mau96]. Hyperwaven tietomalli perustuu hierarkisiin kokoelmiin. Palvelimen juurikokoelmaa lukuunottamatta jokainen dokumentti tai kokoelma kuuluu johonkin kokoelmaan. Kokoelman määritelmä on rekursiivinen siten, että kokoelma voi sisältää toisia kokoelmia. Tämä muodostaa hierarkian kokoelmien välille. Hierarkian täytyy olla syklitön, jotta esimerkiksi kokoelmaan kuuluvien dokumenttien lukumäärä voidaan laskea rekursiivisella algoritmilla. Nämä kokoelman määrittävät säännöt ovat voimassa vain palvelinkohtaisesti.

Kokoelman ulkoasu riippuu selaimesta, jolla kokoelmarakenne tulkitaan. Useimpien kokoelma esitetään käyttäjälle valikon kaltaisena. Käyttäjä valitsee kokoelmahierarkiasta jonkin linkin, joka johtaa hierarkiassa eteenpäin. Kuvassa 1 on eräs kokoelma Harmony-selaimen kautta nähtynä. Harmony sisältää kokoelmaselaimen (vrt. tiedostonhallinta Windows:ssa), jolla kokoelmarakenne mallinnetaan. Kokoelmaselaimen avulla hypertekstin looginen rakenne ilmaistaan käyttäjälle sellaisena kuin tekijä on tarkoittanut. Maurerin mukaan tämä auttaa käyttäjää luomaan hypermedia-aineistosta ajatuksellisen mallin paremmin kuin perinteisessä solmuista ja linkeistä koostuvassa rakenteessa (kuten www). Toisaalta samankaltainen hierarkia voidaan luoda myös www-ympäristössä, mikäli hypertekstin tekijä suunnittelee vastaavanlaisen hypertekstirakenteen ja käyttöliittymän. Www:n linkkimalli ei kuitenkaan sisäänrakennetusti tue tällaista rakennetta, vaikka se onkin mahdollista toteuttaa.

Kokoelmahierarkia ratkaisee osaltaan ongelmaa, jos käyttäjä ei ole selvillä sijainnistaan hypertekstien joukossa. Nykyinen sijainti hypertekstijoukossa ja sen sijainti kokoelmaan käyvät ilmi kokoelmaselaimesta. Kokoelmarakenteen etuna on myös se, että uudet dokumentit tulevat automaattisesti osaksi kokoelmahierarkiaa, eikä uuteen dokumenttiin tarvitse eksplisiittisesti lisätä viitteitä muista dokumenteista.



Kuva 1: Kokoelma Harmony-selaimen tulkitsemana

Hyperwavessa linkit ovat kaksisuuntaisia. Dokumentista nähdään luonnollisesti, mitä ulospäinsuuntautuvia linkkejä se sisältää. Kaksisuuntaisuuden ansiosta voidaan myös selvittää, mistä dokumenteista on linkki tiettyyn dokumenttiin. Linkki voi viitata joko kokonaiseen dokumenttiin tai vain osaan siitä (esim. yhteen kappaleeseen). Kuten edellä on mainittu, linkit tallennetaan omaan rakenteeseensa, joka on dokumenteista erillinen. Tästä on etuna esimerkiksi se, että käyttäjä voi tehdä merkintöjä dokumenttiin, vaikka itse dokumenttiin ei olisi kirjoitusoikeutta. Käyttäjän tekemät muistiinpanomerkinnot nimittäin toteutetaan Hyperwavessa linkkeinä. Lisäksi tallennusratkaisun ansiosta Hyperwave pystyy valvomaan linkkirakenteen eheyttä. Kun jokin dokumentti poistetaan, Hyperwave merkitsee siihen viittaavat linkit avoimiksi. Tämä tarkoittaa sitä, että normaali käyttäjä ei näitä linkkejä sen jälkeen näe, ellei dokumenttia lisätä uudelleen. Hyperwave takaa linkkirakenteen eheyden säilymisen

paikallisesti, eli siinä tapauksessa, kun linkin lähde- ja kohdedokumentti sijaitsevat samalla palvelimella. Hyperwave-kirja tekijät ovat kuitenkin myös esittäneet menetelmän, miten eheyttä voitaisiin valvoa myös globaalisti.

Linkkien ja dokumenttien tallennuksen erottaminen toisistaan mahdollistaa sen, että linkkejä voidaan lisätä videoihin ja kuviin. Itse objekteihin, kuten videotiedostoon, ei tule mitään muutoksia. Linkki lisätään vain Hyperwaven hallintarakenteisiin.

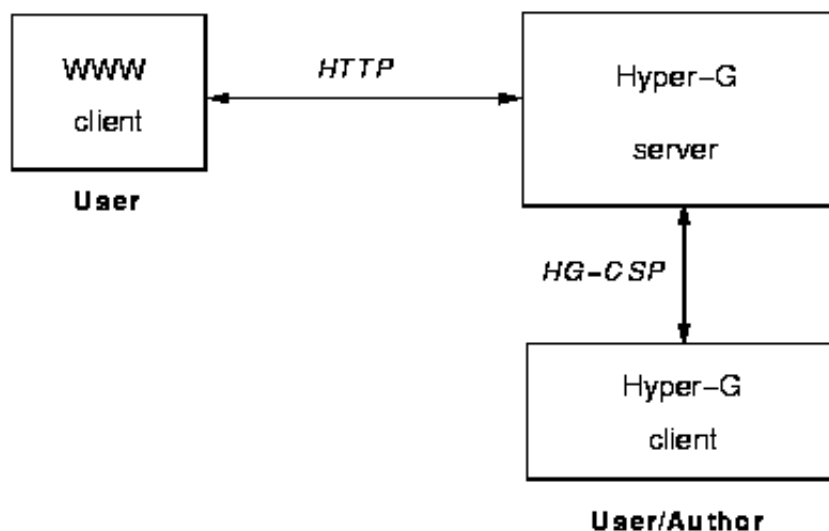
Jokainen objekti tallennetaan tietokantaan, jota Hyperwave-palvelin ylläpitää. Objekteilla, kuten dokumentit ja linkit, on tietty luokka. Dokumentit kuuluvat luokkaan `Document`. Kokoelmat ovat luokan `CollectionDocument` ilmentymiä, joka on `Document`-luokan aliluokka. `Document`-luokka sisältää attribuutteja, kuten `Title`, `DocumentType` ja `Author`, jotka periytyvät myös luokkaan `CollectionDocument`. Koska Hyperwavessa voidaan tehdä hakuja myös attribuutteihin, haut voivat näin ollen palauttaa yhden dokumentin lisäksi myös kokoelmia. Linkit tallennetaan Hyperwavessa lähde- ja kohdedokumenttipareina, ns. ankkureina. Lähdeankkuri sisältää tiedon siitä dokumentin osasta, joka linkin muodostaa. Kohdeankkurissa taas määritellään se osa kohdedokumenttia, johon linkin on tarkoitus osoittaa.

4 Hyperwave-palvelin

Luvussa 4 käsitellään Hyperwaven asiakas-palvelin-mallia sekä palvelimen arkkitehtuuria (Hyperwave-kirjan luvut 4 ja 10 vastaavasti [Mau96]). Lisäksi kuvataan lyhyesti palvelimen tekstihakuominaisuuksia (kirjan luvut 6 ja 10 [Mau96]).

4.1 Asiakas-palvelin -malli

Hyperwave-järjestelmässä asiakasohjelmistojen kautta noudetaan objekteja (esim. dokumentteja, kuvia jne.) Hyperwave-palvelimelta. Kuva 2 esittää tätä rakennetta.



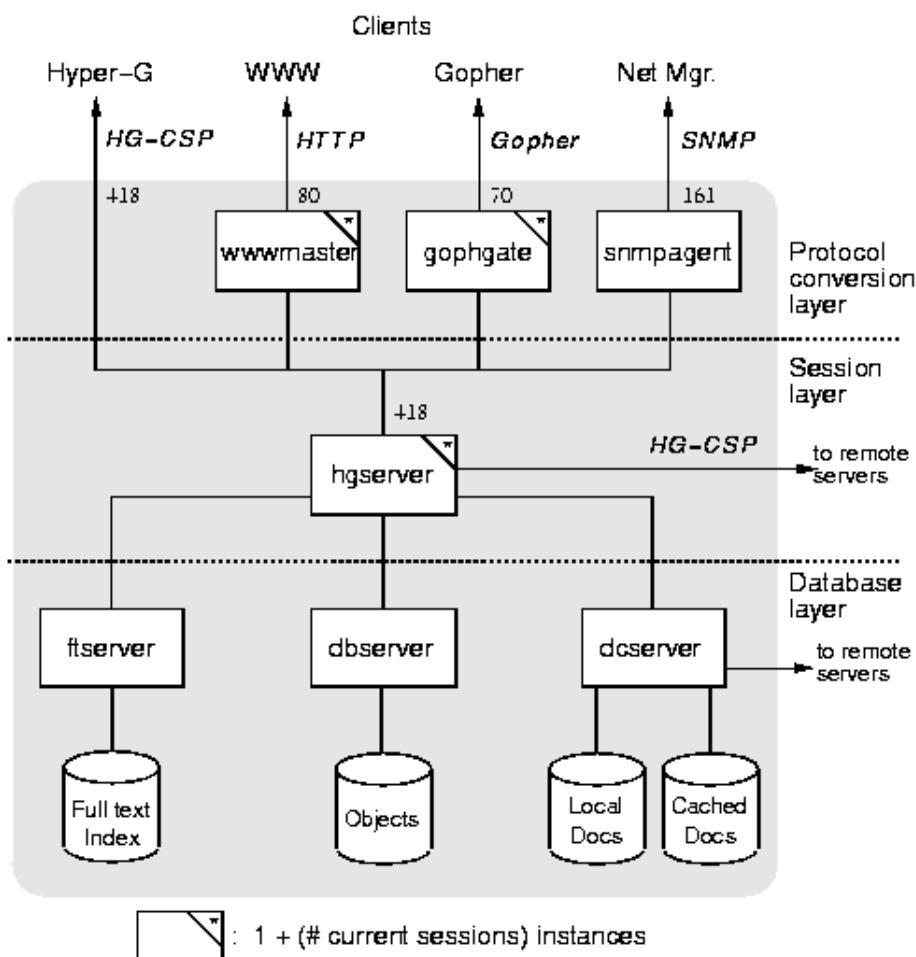
Kuva 2: Hyperwaven asiakas-palvelin-malli

Hyperwave-palvelimelle tallennettuja dokumentteja voidaan käsitellä nykyisillä www-selaimilla. Selain pyytää ja ottaa vastaan hypertekstidokumentit HTTP-protokollan mukaisesti. Hyperwave-palvelin muodostaa HTML-tulkinnan pyydetystä dokumenttijoukosta ja tällöin osa Hyperwaven ominaisuuksista jää hyödyntämättä. Näitä ovat esimerkiksi kokoelmaselaimen käyttö (ks. luku 3) sekä hypermediaympäristön visualisoinnit. Hyperwave-kokoelmat vastaavat www-selaimessa valikoiden kaltaisia valintalistoja. Hyperwaven tekstihaku toteutetaan perinteisenä HTML-kaavakkeena.

Hyperwavea tukevat selaimet, kuten Harmony, kommunikoivat palvelimen kanssa käyttäen Hyperwaven omaa HG-CSP-protokollaa (Hyper-G Client Server Protocol). HTTP-protokollan ollessa yhteydetön (engl. *connectionless*), HG-CSP on yhteydelinen (engl. *connection-oriented*), kuten FTP. HG-CSP-protokollassa yhteys muodostetaan kerran ja sitä pidetään auki session ajan eli kunnes yhteys suljetaan tai tulee aikakatkaistu.

4.2 Palvelimen arkkitehtuuri

Hyperwave-palvelin koostuu monesta moduulista, jotka on toteuttu rinnakkaisina Unix-prosesseina. Moduulit on järjestetty kolmikerrosrakenteeseen. Kuva 3 esittää Hyperwave-palvelimen moduuleja ja kerroksia, joissa moduulit sijaitsevat.



Kuva 3: Hyperwave-palvelimen arkkitehtuuri

Protokollan muunnoskerros (engl. *Protocol conversion layer*) vastaa siitä, miten Hyperwave-dokumenttien sisältö muunnetaan tiettyjen protokollien mukaiseksi. Kerroksessa tehdään muunnos muiden protokollien ja Hyperwaven HG-CSP:n välillä, jotta palvelimen alemmissa kerroksissa voidaan käsitellä vain HG-CSP-protokollaa. Moduuli `wwwmaster` määrittää Hyperwave-palvelimelle tallennettujen dokumenttien

ulkoasun www-selaimia varten. Moduulissa täytyy ratkaista miten esimerkiksi koelmat esitetään HTML-rakenteena.

Sessiokerros (engl. *Session layer*) ottaa vastaan asiakkaalta yhteydenmuodostuspyyntöjä. Yhtä sessiota varten käynnistetään yksi **hgserver**-prosessi. Sessiokerros myös keskustelee muiden Hyperwave-palvelimien kanssa, jos niistä täytyy hakea objekteja. Tietokantakerrokselta (engl. *Database layer*) sessiokerros pyytää tiedonhallintaan liittyviä palveluja, kuten objektien tallettamista tietokantaan.

Tietokantakerros koostuu kolmesta moduulista. Kaikki objektit tallennetaan objektipalvelimelle **dbserver**, joka on monen käyttäjän tietokanta. Tietokannan rinnakkaisuusongelmat on ratkaistu siten, että **hgserver** jakaa pyytämänsä transaktion osiin, jotta **dbserver** voi käsitellä rinnakkain pyyntöjä eri asiakkailta. Tietokantakerroksen transaktiot suoritetaan sarjallisesti, jolloin vältetään se, että kaksi transaktiota muokkasi samaa objektia yhtä aikaa. Hyperwaven välimuistipalvelin on **dcserver** (*Document cache server*). Välimuistipalvelin noutaa asiakkaan pyytämiä dokumentteja sekä tallettaa niitä omaan tietokantaansa. Lisäksi palvelin ottaa vastaan dokumentteja muilta Hyperwave-palvelimilta ja tallentaa ne välimuistiin. Välimuistipalvelin ajaa lisäksi CGI-skriptit ja palauttaa niiden tuloksen käyttäjälle. **Ftserver**-palvelin (*Full text server*) tallentaa dokumenttien termeistä muodostetun käänteisindeksin. Kun dokumentti lisätään Hyperwaven tietokantaan se lähetetään heti **ftserver**:lle, joka purkaa dokumentista termit ja lisää ne käänteisindeksiin. Käänteisindeksi on rakenne, jossa jokaista termiä vastaa lista niiden dokumenttien tunnuksista, jotka sisältävät kyseisen termin [Ara01]. Lisäksi dokumentin tunnuksen yhteydessä on tieto termin sijainnista dokumentissa.

4.3 Tekstihaku aineistosta

Kuten johdannossa mainittiin, Hyperwaven tekijät pitävät sisäänrakennetun tekstihakujärjestelmän puuttumista yhtenä `www:n` haittapuolista. Hyperwave-palvelimen tietokannasta voidaan hakea tekstitietoa mistä tahansa sinne talletetusta objektista. Objektien attribuutit voivat myös olla hakujen kohteena. Tietyt attribuutit, kuten `Author` ja `Title`, ovat indeksoituja hakujen nopeuttamiseksi. Tietoa voidaan hakea koko Hyperwave-palvelimen sisällöstä tai sitten haku voidaan rajata koskemaan joitain tiettyjä kokoelmia. Haut eivät ole siinä mielessä rajoittuneita yhdelle Hyperwave-palvelimelle, että kokoelma voi sijaita osittain myös muilla palvelimilla. Hyperwavessa hakutermeistä voidaan muodostaa Boolean lausekkeita yhdistelemällä niitä konnektiiveilla `AND`, `OR` ja `AND NOT`. Hakutermien tärkeyttä haussa voidaan painottaa eri tavoilla. Lisäksi hauissa voidaan käyttää vain osaa sanasta. Hyperwave-palvelin tukee listaa yleisistä sanoista, joita ei oteta haussa huomioon (engl. *stop words*). Lisäksi hakutermejä voidaan supistaa (engl. *stemming*) katkaisemalla niistä esimerkiksi sijapäätteitä pois. Tekstihaun tulosjoukon Hyperwave muodostaa hyväksikäyttämällä luvussa 4.2 esiteltyä `Ftserver`-palvelimen muodostamaa käänteisindeksiä.

5 Hyperwave nykyään

Hyperwave ei tätä nykyä mainosta itseään niinkään `www`-palvelimena, vaan tiedonhallintajärjestelmänä ja organisaation yhteistyön kehittämisen välineenä [Hyp04].

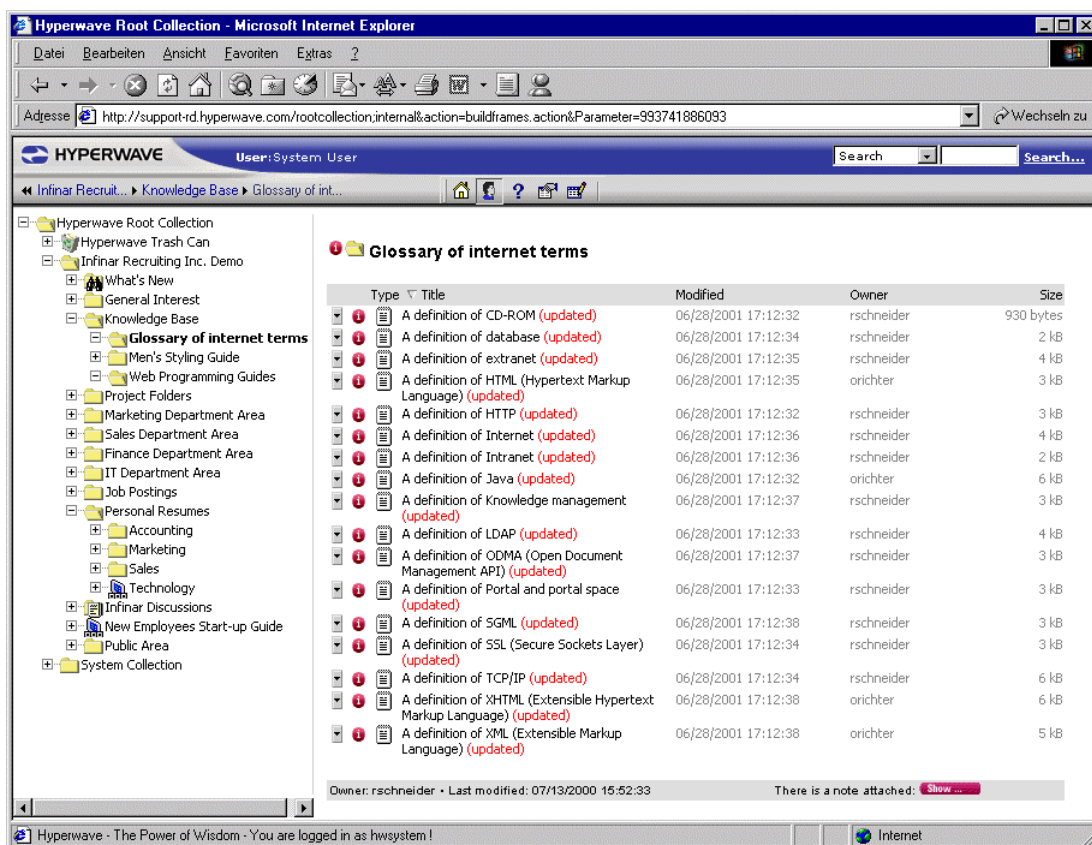
Kaikkien Hyperwave-tuotteiden pohjalla on Hyperwave IS/6 (ent. Hyperwave Information Server) [Kap01b]. Hyperwave IS/6 mahdollistaa Hyperwaven tiedonhallintaa hyväksikäyttävien verkkopohjaisten sovellusten teon. IS/6 sisältää kaiken sitä käytävien sovellusten ydintoiminnallisuuden. Näitä ovat mm. dokumenttien linkkiraken-

teen ylläpito, tekstihakujärjestelmä sekä dokumenttien versionhallinta. Hyperwave IS/6 ei enää sisällä omaa tietokantaa vaan se vaatii joko Oraclen tietokannanhallintajärjestelmän tai Microsoftin SQL Serverin. Hyperwaven oma protokolla HG-CSP ei myöskään ole enää käytössä, vaan kommunikaatio Internet-selaimen välillä tapahtuu HTTP-protokollalla.

Yksi Hyperwaven omista IS/6:n päällä toimivista sovelluksista on Hyperwave eKnowledge Suite. Kuvassa 4 on esimerkki sovelluksen käyttöliittymästä. Järjestelmä on tarkoitettu erityisesti dokumenttien ja tiedon hallintaan. Se sisältää mm. dokumenttien linkkirakenteen eheyden ylläpidon ja sisäänrakennetun hakujärjestelmän. Lisäksi järjestelmä tukee dokumenttien versionhallintaa ja automaattista luokittelua. Mukana on myös erilaisia kommunikointivälineitä, kuten tuki keskustelufoorumeille ja yhteinen työskentelytila hajautetuille tiimeille. Hyperwave eLearning Suite taas on verkkopohjaiseen opiskeluun suunniteltu tuote.

6 Yhteenveto

Hyperwave-kirjassa [Mau96] vuonna 1996 kuvattu Hyperwave-järjestelmä on nykyään kaupallisessa käytössä, mutta se ei ole kuitenkaan pystynyt nousemaan perinteisten www-palvelimien haastajaksi. Tähän on vaikea löytää yksiselitteistä syytä. Hyperwaven ja www:n alkuvaiheen kehitys tapahtui melko samaan aikaan. Www sai kuitenkin alussa valta-aseman Mosaic-selaimen myötä, joka otettiin hyvin vastaan ja tuli tiedeyhteisön käyttöön. Maurer toteaaakin joutuneensa tekemään Hyperwaves-ta vahvasti yhteensopivan HTTP:n ja HTML:n kanssa osittain juuri sen takia, että www oli jo saanut niin vahvan aseman. Hyperwave on myös hierarkisempi tiedonhallintajärjestelmä kuin www, joka ei sinänsä pakota mihinkään hierarkiaan linkkien ja solmujen välillä. Hyperwavessa tieto järjestetään kokoelmiksi, jotka sisältävät alikokoelmia, hieman kuin tiedostonhallinta esim. Windows:ssa ja Unix:ssa. Voi olla,



Kuva 4: Hyperwave eKnowledge Suite ([Kap01a])

että ww:n salliva hypermediamalli miellyttää ihmisiä enemmän kuin tiettyyn rakenteeseen pakottava ratkaisu.

Lähteet

- Ara01 Arasu, A. et al., Searching the web. ACM Transactions on Internet technology, Vol. 1, No. 1 (2001), 2-43.
- Goo04 Google-hakukone, www.google.com. [25.11.2004]
- Hyp04 Hyperwave, www.hyperwave.com. [25.11.2004]
- Kap01a Kappe, F., Hyperwave eKnowledge Infrastructure. Technical white paper, 2001.
- Kap01b Kappe, F., Knowledge Management with the Hyperwave eKnowledge Infrastructure. Technical white paper, 2001.
- Mau96 Maurer, H., Hyperwave - The Next Generation Web Solution. Addison Wesley Longman, Harlow, 1996.