

'vapaamuotoiset' tiedot

- Relaatiotietokantaan tallennetaan määrämuotoista tietoa
- Vapaamuotoisen tekstin talletukseen **tekstitietokannat**
 - hakuindeksejä
 - saattavat käyttää relaatiotietokantaa tekstin tallentamiseen rivipohjainen tai dokumenttipohjainen tallennus

'vapaamuotoiset' tiedot

- Vähemmän 'vapaamuotoisen' materiaalin esitysmuotoja
- HTML
 - dokumentit tallennettu tiedostojärjestelmään
 - dokumentit kytkettävissä yhteen **hyperlinkkien** avulla
 - HTML -kieli määrittelee dokumentin teknisen rakenteen (otsakkeet, kappaleet, taulukot,...)
 - ulkoasun pääpiirteet - selain päättää yksityiskohdat

'vapaamuotoiset' tiedot

- HTML on rakenteisten dokumenttien määrittelyyn tarkoitettu SGML-standardin sovellus
- SGML-on merkkaukielimäärittely
 - Dokumentin elementit merkataan:
 - alkumerkki - loppumerkki
 - <element> elementti </element>
- HTML:ssä elementit rakenne-elementtejä
- <h1>Tason 1 otsake</h1>
- <table><td>taulukkoalkio</td> ... </table>

'vapaamuotoiset' tiedot

- HTML-dokumenttia luetaan selaimella (browser)
- HTML-dokumentti voi nykyään sisältää tekstiä, kuvia, ääntä, videota, suoritettavia ohjelmia
- HTML-dokumentti on
 - luettavissa missä tahansa ei sidoksia tiettyyn käyttöympäristöön
 - tallennettavissa minne tahansa
 - Internet/intranet -pääsy
 - Internetin kautta pääsy miljardeihin HTML-dokumentteihin - valtava dokumenttikanta

'vapaamuotoiset' tiedot

- Ongelmana tieton löytäminen valtavasta dokumenttikannasta
- HTML-dokumenttikantaan ei voi soveltaa SQL-tyyppistä hakua

```
select tieto from dokumenttikanta where asia='haluttu'
```
- sillä HTML:ssä ei asianmukaista merkintää.
- HTML-dokumenttikantaan on rakennettu tekstitietokantatyyppisiä hakuindeksejä esim. Altavista, Google, HotBot

'vapaamuotoiset' tiedot

- Hakuindeksien toiminta vastaa lähinnä SQL-kyselyä

```
select URL from dokumenttikanta where dokumentti like '%hakusana%';
```
- Saatavia dokumentteja on yleensä hyvin paljon ja suurin osa ei ole sitä mitä etsitään
- Eri hakukoneet järjestävät tuloksen eri tavoin
 - ei järjestystä,
 - uusimmat ensin,
 - eniten viitatu ensin

'vapaamuotoiset' tiedot

- XML on asianperusteiseen merkkauseen perustuvan standardi - 'SGML:n yksinkertaistus'
- Tietylle dokumenttityypille, vaikkapa rakennussuunnitelmalle, voidaan määrittellä oma XML-standardia noudattava esityskielensä. Rakennussuunnitelmakielessä (olkoon se REXML) voisi olla elementtejä kuten <yleispiirustus>, <sähkösuunnitelma>, <huone>, <katto>, <seinä>, <ikkuna>, jne

'vapaamuotoiset' tiedot

- Perus XML ei määrittele esitysmuotoa kuten HTML, vaan esitysmuotoja voidaan määrittellä erillisen esitysmäärittelyn (tyylitiedoston) avulla. Samalle dokumentille tai pikemminkin dokumenttityypille voidaan laatia useita erilaisia esitysmäärittelyitä.
- XML-selain osaisi esitysmäärittelyjen perusteella näyttää dokumentin.

'vapaamuotoiset' tiedot

- XML mahdollistaa asiapohjaisen haun

```
select URL from REXML-dokumenttikanta  
where 'pelti' in rakennussuunnitelma.katto.materiaali
```
- Tällaiset haut edellyttävät monipuolisempia hakuindeksejä

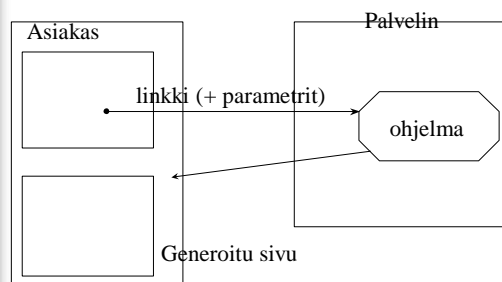
WWW ja tietokannat

- WWW-sivut
 - HTML-kielillä esitettyä hypertekstiaineistoa
 - tekstiä, kuvia, hyperlinkkejä
- Staattiset sivut
 - kirjoitettu kerran, muuttaminen käsin
 - ongelmana pysyminen ajantasalla
 - Ylläpito hankalaa, jos sama tieto esiintyy usealla sivulla
 - Tiedon löytäminen vaikeaa, jollei sisäistä sivun laatijan käyttämää 'jäsenysmallia'. Jäsentäminen eri tavoin johtaa ylläpito-ongelmiin

WWW ja tietokannat

- Ylläpitoa helpottamaan: **generoidut sivut**
 - Sivut generoidaan ohjelmallisesti tietokantaan sisältyvän aineiston perusteella
 - ajoittain
 - esimerkiksi aina kun dataa on muutettu
 - kerran viikossa
 - aina pyydettyäessä (=kun joku pyytää sivua)
- Pyydettyäessä laadittavat dokumentit rakennetaan erillisten palvelimissa toimivien ohjelmien avulla.
- CGI (common gateway interface)

WWW sivun generointi



CGI

- CGI:n kautta käynnistettävälle ohjelmalle voi välittää parametreja
 - kirjoittamalla ne URL:iin
 - teos.haku?isbn=0-123456-67-1&kkoodi=FIN
 - antamalla ne lomakkeella
 - ohjelma saa parametrit standardisyöttövirrassa tai ympäristömuuttujan kautta
- palvelimessa toimiva ohjelma voi tehdä mitä tahansa (kirjata tilauksen, tuottaa uuden www-dokumentin, laskuttaa 1Mmk asiakkaan luottokortilta,...)

CGI

- CGI-ohjelmat muodostuvat käynnistyessään aina oman prosessinsa, jonka ympäristö (tilatiedot) katoaa prosessin päätyessä
- CGI-ohjelma ei voi säilyttää yhteyskohtaista tietoa, joka välittyisi kutsukerrasta toiseen
 - tällöin esimerkiksi CGI-ohjelman luomat tietokantayhteydet katoavat ja ne joudutaan perustamaan uudelleen seuraavan kutsun yhteydessä

CGI

- CGI:lle vaihtoehtona on esitelty Java Servlet -tekniikkaa, jossa palvelinprosessi onkin sitkeä ja jää henkiin ja uudelleentavoitettavaksi suoritettuaan palvelun, mahdollistaa esim. tietokantayhteyksien säilyttämisen, jolloin palvelu nopeutuu.
- Java Servlet tekniikassa parametrit välittyvät CGI-tekniikalla
- Jotkin WWW-palvelimet osaavat myös säilyttää prosessin tilan ja tietokantayhteydet omilla tekniikoillaan

Lomake -WWW sovelluksen perusta

```
<FORM miten ja minne>  
lomakkeen kenttien määrittely  
</FORM>
```

```
<FORM method="POST" ACTION=  
"HTTP://www.helsinki.fi/cgi-bin/ohjelma">  
lähettää lomakkeen tiedot ohjelma -nimiselle ohjelmalle  
hakemistossa cgi-bin
```

Lomake -WWW sovelluksen perusta

- `<INPUT TYPE="tyyppi"
NAME="nimi" VALUE="arvo">`
- tyyppi:
 - **text** - tekstikenttä
 - **radio** - poissulkeva valintanappi (name kertoo ryhmän, johon nappi kuuluu)
 - **checkbox** - valintanappi

Lomake -WWW sovelluksen perusta

- lisää tyyppejä:
 - **submit** - lähetyksenappi
 - **reset** - tyhjennysnappi
 - **password** - kuten text, mutta syötettävä teksti ei näy
 - **hidden** - kokonaan piilossa oleva kenttä (näky source-muodossa)

Lomake -WWW sovelluksen perusta

Luettelo ryhmään ilmoittautuneista

Kurssin nimi:

Ryhmän numero: Henkilötunnuksilla:

Lomake -WWW sovelluksen perusta

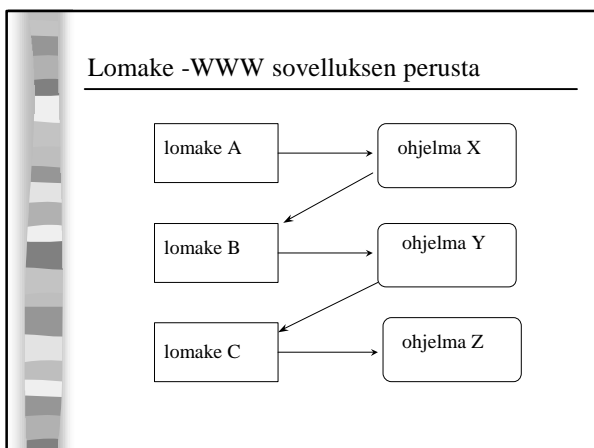
```
<hr><H2> Luettelo ryhmään ilmoittautuneista</H2>
<Form Method="Get" Action= ohjelman_URL>
Kurssin nimi: <input type="text" name="kurssi" size= 40
maxlength=40> <p>
Ryhmän numero: <input type="text" name="ryhma" value=""
size=4 maxlength=4>
Henkilötunnuksilla: <input type="checkbox" name="cb1"
value=" tunnuksin" ><p>
<input type="submit" value="Tee lista">
<input type="reset" value="Tyhjennä">
</form><hr>
```

Lomake -WWW sovelluksen perusta

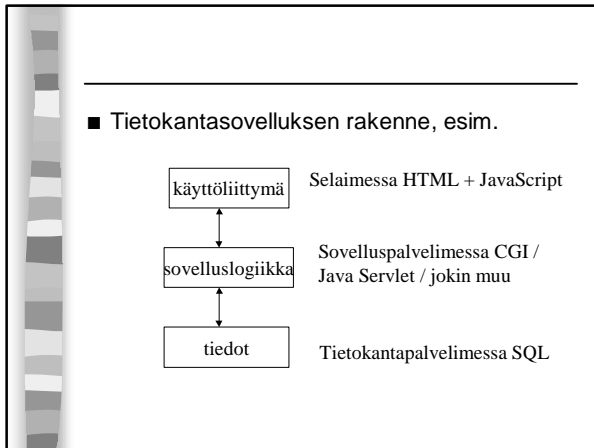
■ Valikot

```
<Select name="valikko" size=5>
<option selected>Ensimmäinen
<option>Toinen
....
<option>Kahdeskymmenes
</select>
```

- Lomake -WWW sovelluksen perusta
- Lomakkeen elementtejä vastaavat parametrit lähetetään ohjelmalle nimi=arvo -pareina yleensä siinä järjetyksessä kuin kentät on lomakkeella määritelty
 - Lomake voi olla jonkin ohjelman generoima, näin voidaan rakentaa useasta lomakkeesta muodostuva käyttöliittymä



- Lomake -WWW sovelluksen perusta
- Tarkistukset tehtävä palvelimessa
 - Tarkistukset voi tehdä myös selaimessa jos käytävissä on jokin skriptikieli esim. Javascript
 - Sovelluslogiikkaa asiakaskoneeseen
 - Aktiiviset komponentit
 - Java sovelmat (applet)
 - Javascript tarkistukset
 - Active X, yms.



WWW-pohjaisista tietokantasovelluksista

- Palvelimessa suoritettavat
 - CGI
 - Servlet:it
 - JavaScript
 - ASP(Active Server Pages - Microsoft), JSP
 - PHP (Personal Home Pages)
- Selaimessa suoritettavat
 - JavaScript, VBScript
 - Java sovelmat (Applet)
 - Active X -komponentit

Tietokantasovelluksen toteutustapoja

- Vaihtoehtoja:
- cgi-skripti ->tkhj
 - ongelma tietokantayhteyden luominen aina uudelleen, jos samaan kantaan useita operaatioita
 - voidaan tietenkin rakentaa erillinen yhteysmanageri, joka pitää yhteydet hengissä
- java servlet: -> tkhj
 - yhteys voidaan säilyttää
- erityinen tietokantaa hyödyntävä www-palvelin, esim. Oracle Web Server

Tietokantasovelluksen toteutustapoja

- Java-sovelma tai ActiveX-komponentti
 - toimii selaimessa ja kommunikoi suoraan tietokantaohjelmiston kanssa

Oracle Web Server

- pystyy käynnistämään CGI:n ohi suoraan tietokantaproseduureja ja välittämään niille kutsussa saadut parametrit valmiiksi purettuina proseduurin parametreiksi (yleensä CGI-ohjelmassa on erikseen purettava parametrit)
- tietokantayhteyden kuvauksessa määritellään, millä käyttäjätunnuksella tietokantaproseduuuri suoritetaan
- URL:ssä nimetään käytettävä tietokantayhteys

Oracle Web Server

- Kannankäsittely tietokantaproseduureilla
- Tietokantaproseduureja laadittaessa on käytettävissä erityiset ohjelmakirjastot html:n tuottamiseen
 - http-kirjasto sisältää tulostusproseduurin jokaista html-merkintää varten, sekä yleisen tulostusproseduurin print
 - http.print("Tulostettava teksti");

- Java Servlet ratkaisussa servletiluokat perivät luokan HttpServlet. Tarjoaa palvelut
 - parametrien arvojen purkamiseen
 - tulostuksen ohjaukseen selaimelle
- Servlettejä ajetaan servlettiympäristössä (java virtualikone + kytkentä www-palvelimeen)
- Servletti toteutetaan luokkana. Servlettiympäristö lataa luokan vain kerran ja luo sille yhden ilmentymän. Palvelupyynnöt hoitavaa metodia ajetaan säikeinä.

