CDL 2005 kokosi digitaalisen kirjaston kehittäjät

Tietojenkäsittelyn ja kirjastoalan välimaastoon sijoittuva ECDL 2005 -konferenssi järjestettiin tänä syksynä yhdeksättä kertaa, tällä kertaa Wienissä 18-23. syyskuuta. Tämä konferenssisarja on vakiinnuttanut asemansa keskeisenä eurooppalaisena digitaalisen kirjaston tutkijoiden ja kehittäjien foorumina, mikä kuvastui myös osanottajien määrässä (n. 400), joista suomalaisia osallistujia oli vain viisi henkeä.

Konferenssin koko nimi oli ***9th European Conference on Research and Advanced Technology for Digital Libraries***. Konferenssi oli tutkimuspainotteinen, mutta lähestymistavaltaan hyvin laaja-alainen, esitelmien teemat vaihtelivat FinELibin käyttäjätutkimuksesta henkilökohtaisiin digitaalisiin arkistoihin ja tulevaisuuden vertaisverkkoihin pohjautuvaan digitaalisten kirjastojen arkkitehtuuriin.

Konferenssipaikka oli Freihaus-niminen Wienin teknillisen yliopiston rakennus. Tällä paikalla oli pari sataa vuotta sitten samanniminen rakennus, jossa Mozart sävelsi Taikahuilun.

Ennen varsinaisen konferenssin alkua oli tarjolla aihepiiriin johdattelevia tutoriaaleja. Osallistuin uusiseelantilaisen **Ian Wittenin** vetämään osioon, joka käsitteli digitaalisen kirjaston käytännön yhteentoimivuusstandardeja.

Lähestymistapa oli käytännönläheinen ja demoihin perustuva, esimerkiksi osanottajilla oli mahdollisuus asentaa omalle kannettavalleen avoimeen lähdekoodiin perustuva Greenstone-ohjelmisto, jota käytettiin esimerkkinä.

Esimerkkien avulla käsiteltiin sellaisia standardeja kuin MARC (Machine-Readable Cataloging), Z39.50, Dublin Core, METS, MODS, OAI-PMH ja SRW/U. Erityisesti MODS-standardista mainittiin, että tämä Kongressin kirjaston kehittämä metadatastandardi tulee jatkossa löytymään monesta paikasta, sitä kuvailtiin yhdistelmäksi Dublin Corea ja MARCia. SRW/U-tiedonhakustandardia luonnehdittiin paluuksi webpohjaiseen, yleisesti käytettyyn tekniikkaan kirjastoalan omien erikoislaatuisten standardien jälkeen (MARC ja Z39.50). SRW/U-standardille löytyy käyttöä myös Nelli-portaalissa, jossa sen avulla voidaan integroida esimerkiksi DSpace-pohjaiset palvelut Nellissä haettavaksi kohteeksi.

Greenstone-ohjelmisto vaikutti hyvin monipuoliselta ja monia standardeja ymmärtävältä alustalta, se esimerkiksi osaa siirtää tiedot DSpace-ohjelmiston käyttämässä formaatissa. Greenstone-ohjelmistoa kehitetään Waikaton yliopistossa Uudessa Seelannissa, ja sillä on yhteistyötä Unescon kanssa. Se toimii alustana myös New Zealand Digital Library -projektissa. Waikaton yliopisto sijaitsee surffaajien suosimalla alueella, sillä siellä on tiettävästi parhaat vasemmalta kaartuvat aallot surffausta silmälläpitäen.

Henkilökohtaiset digitaaliset arkistot

Konferenssin ensimmäisen keynote-esitelmän piti **Neil Beagrie** British Librarysta aiheenaan henkilökohtaiset digitaaliset arkistot. Perinteisesti esimerkiksi kirjailijoiden henkilökohtaiset arkistot ja kirjeenvaihto ovat olleet kirjallisuuden ja historian tutkijoille tärkeää aineistoa. Kuinka moni nykykirjailija tallentaa systemaattisesti kaikki lähettämänsä ja vastaanottamansa sähköpostit? Esimerkkinä kirjailijasta, joka tämän tekee mainittiin Salman Rushdie.

Tulevaisuudessa kirjastot ovat mielenkiintoisen haasteen edessä, kun niille lahjoitetaan kulttuurihenkilöiden arkistoja sähköisessä muodossa, esimerkiksi Nokia Lifeblog -formaatissa tai MyLifeBits-hankkeen kokeileman jatkuvan videotallennuksen tuotoksena. Tähän aihepiiriin liittyvät myös opiskelijoiden tuottamat sähköiset portfoliot, joiden tulisi mielellään säilyä yhtä lukukautta pidempään eräänlaisina työnäytteinä, joille voi olla käyttöä myöhemminkin.

Muistiin ja muistamiseen liittyy myös Memories for Life -hanke, jossa aihetta lähestytään monitieteisesti neurotieteen, psykologian ja tietojenkäsittelyn näkökulmista. Edellämäinittu hanke lähtee tutkimuksen näkökulmasta, sen sijaan ourmedia.org-hanke perustuu ruohonjuuritason yhteistyöhön yksittäisten ihmisten kesken. Se lupaa ilmaista säilytystilaa hamaan tulevaisuuteen saakka kaikenlaiselle digitaaliselle aineistolle, esimerkiksi omille videotallenteille. Ainoa edellytys on, että aineistot ovat vapaasti käytettävissä. Esimerkkinä henkilökohtaisten digitaalisten arkistojen näköpiirissä olevista käyttökohteista Beagrie mainitsi sukututkimuksen.

Osa konferenssin esitelmistä olisi varsin teknisiä ja tietojenkäsittelyn matemaattisiin malleihin perustuvia tutkielmia, esimerkiksi **Christos Tryphonopouloksen** esitys LibraRing-arkkitehtuurista. Tutkimuksessa osoitettiin kokeellisesti, että vertaisverkkoihin perustuva P2P-arkkitehtuuri skaalautui, vaikka solmujen määrää lisättiin 100 000:een ja dokumenttien kokoa kasvatettiin merkittävästi. Käytännössä tällaiset tutkimukset antavat viitteitä siihen, että MP3-musiikkitiedostojen jakelussa tehokkuutensa osoittanut vertaisverkko on lupaava malli myös digitaalisten kirjastojen yhteistyöhön. Tähän tavoitteeseen tähtääviä hankkeita onkin käynnissä monella taholla, esim. Delos-verkoston puitteissa.

Käyttäjälähtöisiä tutkimuksia

Vastapainoksi teknisesti painottuneille esityksille kuultiin joukko käyttäjätutkimuksesta lähteviä esitelmiä. **Pertti Vakkari** selosti FinELibin käyttäjäkyselyn aineistoon perustuvaa tutkimusta, jossa selvitettiin tieteenalakohtaista vaihtelua tiedonlähteiden käytössä. Tarkasteltaessa kaikkia tieteenaloja tiedonlähteiden tärkeysjärjestys oli: 1. haku lehtitietokannoista 2. haku viitetietokannoista 3. sisällysluetteloiden selaus 4. kirjallisuusviitteiden seuraaminen 5. kysyminen kollegoilta ja 6. uutuusseurantapalvelut.

Toinen esimerkki käyttäjälähtöisestä tutkimuksesta oli **Frank McCown** esittelemä tiedonlähteiden evaluointi, jossa 11 amerikkalaista keskiasteen opettajaa arvioi ensin sopivimmat hakutermit ja sen jälkeen näillä löydetyt hakutulokset. Tulokset olivat peräisin joko laajasta NSDL-palvelusta (National Science Digital Library) tai Googlesta, mutta opettajat eivät tienneet hakutulosten alkuperää, vaan arvioivat niiden käytettävyyttä oman opetuksensa kannalta. Tässä vertailussa Googlen kautta löydetyt aineistot arvioitiin käyttökelpoisemmiksi ja Googlen tapa järjestää hakutulokset vastasi paremmin asiantuntijoiden arvioita kuin NSDL-palvelusta saadut tulokset. Olisi kiinnostavaa nähdä mikä olisi tulos, jos vastaava evaluointi tehtäisiin esimerkiksi Google Scholarin ja ScienceDirectin kesken.

Digitaalisen arkistoinnin kysymykset

Konferenssiin sisältyi oma sessio digitaalisen arkistoinnin kysymyksille. Kiinnostavan esityksen DSpace-ohjelmistosta piti **Robert Tansley**, joka on yksi hankkeen pääsuunnittelijoista. Hänen näkökulmansa oli kuvata miten DSpace-ohjelmiston kehitys avoimen lähdekoodin hankkeena on lähtenyt käyntiin. Avoimen lähdekoodin hankkeissa keskeistä on, että käyttäjäyhteisö alkaa aktiivisesti tuottaa parannuksia havaittuihin ongelmiin ja kehittää sellaisia uusia piirteitä, joita käyttäjät pitivät tarpeellisina. DSpace on Tansleyn mukaan ylittänyt tämän jatkokehityksen kannalta kriittisen kynnyksen, sillä sitä kehitetään niin monella taholla, että tämän digitaalisen kirjaston ohjelmiston tulevaisuus näyttää olevan vakaalla pohjalla.

Esimerkkejä eri yliopistoissa, ohjelmiston alkuperäisten kehittäjien ulkopuolella tuotetuista parannuksista ovat kokoelmakohtaisesti räätälöitävät tallennuslomakkeet, kuva-aineiston pienet automaattisesti tuotetut selailukuvat, käyttäjien omien kommenttien lisäysmahdollisuus, tuki LDAP-autentikoinnille

ja monikielisille käyttöliittymille sekä tuki RSS-syötteille. DSpace-ohjelmistolla on varsin laaja käyttäjäkunta myös sellaisissa maissa kuin Kiina ja Intia, joten luultavaa on, että jatkossa osa ohjelmiston uusista piirteistä on peräisin näiden maiden yliopistoista ja tutkimuslaitoksista.

Posterinäyttely

Konferenssiin kuului myös laaja posterinäyttely, ja postereiden esittelyä varten niiden tekijöillä oli jokaisella 90 sekuntia aikaa esitellää omaa projektiaan. Näinkin lyhyen esittelyn perusteella sai vaikutelman siitä onko aihe kiinnostava omalta kannalta. Huomiotani kiinnitti **Elizabeth Liddyn** hanke, jossa tutkittiin oppimateriaalin metadatan automaattista luomista dokumenteista ja verrattiin tulosta manuaaliseen indeksointiin. Näillä kahdella tavalla kuvailtuja dokumentteja käytettiin aineistona tiedonhakukokeessa, jossa koehenkilöt etsivät aiheeseensa sopivia aineistoja. Tuloksena oli, että automaattisesti tuotetut kuvailut ja asiasanat toimivat tiedonhaun kannalta yhtä hyvin kuin manuaalisesti annetut. Tämä tutkimus kumoaa osaltaan yleistä olettamusta, että käsin tehtävä sisällönkuvaus olisi jotenkin korvaamaton ja ainutlaatuinen lisäarvo.

Työpajatyöskentelyä

Konferenssin jälkeen osallistuin NKOS-työpajaan, jossa esiteltiin lähinnä erilaisia tesaurus- ja ontologiaprojekteja, mm. FAO:n Agrovoc-sanaston sovittamista yhteen 60.000 termiä sisältävän kiinalaisen maatalousalan asiasanaston kanssa. Esittelyssä oli myös Tampereella tehty tutkimus verkkoklinikka.fi-palvelussa esitetyistä sydän- ja verisuonitauteja koskevista kysymyksistä ja vastauksista. Noin 65 % kysymyksissä ja vastauksissa käytetyistä käsitteistä löytyi FinMESH-asiasanastosta ja 35-40 % löytyi Yleisestä suomalaisesta asiasanastosta. Ontologiahankkeiden esittely sattui sopivaan ajankohtaan, sillä maatalousmuseo Sarka ja Viikin tiedekirjasto ovat yhteistyössä FinnONTO-hankkeen kanssa aloittamassa maatalousalan MATO-ontologian rakentamista kirjaston ylläpitämän Agriforest-sanaston pohjalta.

Laaja-alaisena foorumina ECDL avasi runsaasti erilaisia lähestymistapoja digitaalisten kirjastojen tutkimukseen ja käytännön kehitystyöhön, josta edellä on tuotu esiin joitakin vaikutelmia. Lisätietoja löytyy konferenssijulkaisusta, joka on saatavilla Lecture Notes in Computer Science -sarjassa:

http://www.springerlink.com/openurl.asp?genre=issue&issn=0302-9743&volume=3652

Kimmo Koskinen
kehityspäällikkö
Viikin tiedekirjasto
puh. 191 58782
kimmo.koskinen at helsinki.fi