

Tämä on rinnakkaistallenne.

Rinnakkaistallenteen sivuasettelut ja typografiset yksityiskohdat *saattavat poiketa* alkuperäisestä julkaisusta.

Julkaisun tekijä(t): Tolonen, Tiina

Julkaisun nimi: Keisarin uudet vaatteet eli Theseuksen DSpace-päivitys kesäkuussa 2017

Julkaisuvuosi: 2017

Versio: Julkaistu versio

Käytä viittauksessa alkuperäistä lähdettä:

Tolonen, T. (2017). Keisarin uudet vaatteet eli Theseuksen DSpace-päivitys kesäkuussa 2017. Kreodi, 2.

20.09.2017, 19:33

KEISARIN UUDET VAATTEET ELI THESEUKSEN DSPACE-PÄIVITYS KESÄKUUSSA 2017

THESEUS

Tiina Tolonen

Kaikkien Kansalliskirjaston ylläpitämien julkaisuarkistoinstanssien taustalla oleva ohjelmisto DSpace päivitetään versioon 5.6 julkaisuarkisto kerrallaan vuonna 2017. Pääosa päivityksistä suunniteltiin tehtäväksi jo ennen kesälomakauden alkamista ja viimeiset elo-syyskuun aikana. Ensimmäisenä päivitettäväksi tuli Luonnonvarakeskuksen julkaisuarkisto Jukuri, ja Theseuksen vuoro oli kesäkuussa.

Mikä on DSpace?

DSpace on julkaisuarkistokäyttöön suunniteltu avoimen lähdekoodin ohjelmisto, joka on ollut erilaisten instituutioiden käytössä ympäri maailman vuodesta 2002 lähtien, Kansalliskirjastossa päätös sen käyttöönotosta tehtiin keväällä 2006. Suomalaisista julkaisuarkistoista ainoastaan Oulun yliopiston Jultika toimii muulla kuin DSpace-ohjelmistolla.

DSpacen suosio perustuu helppoon käyttöönottoon, sen voi ottaa käyttöön ilman mitään sopimuksia tai maksuja. Sitä pidetään luotettavana mutta myös vaikeasti muokattavana. Erilaiset muutokset sekä toiminnallisuuksien lisääminen saattavat hidastaa uusien versiopäivitysten käyttöönottoa ja aiheuttaa niihin bugeja. Tämä lienee osasyynä siihen, että monet amk-rintamalta tulleet Theseukseen liittyvät kehittämissuhteet ovat odotelleet mahdollisuutta toteuttamiseen jo hyvinkin pitkän aikaa, esimerkkinä vaikka YAMK-töiden parempi näkyvyys. DSpacen käyttöön, käyttäjiin ja kehittäjiin voi tutustua tarkemmin osoitteessa dspace.org.

Uutta, vanhaa, sinistä vai vaihdettua?

Theseuksen DSpace-päivitys toi uutta sekä satunnaiselle käyttäjälle että opinnäytetöitä käsittelevälle kirjastohenkilökunnalle. AMKIT-wikin eli eDuunin Theseus-alasivun [yleisohjeista](#) löytyy Minna Marjamaan laatima dokumentti, johon on tarkemmin koottu kirjastohenkilökuntaa varten päivityksen keskeisimmät ja näkyvimmat muutokset. Tässä artikkelissani keskityn enemmän päivityksen taustoihin ja seurauksiin. Päivitetystä Theseuksesta löytyy vielä toimintoja yms. joiden käyttämisestä tai tarkoituksesta ei tietoja em. eDuunin dokumentista löydy. Näihin Kansalliskirjasto on luvannut palata tarkemmin kunhan päivitysruuhka ja sen mukanaan tuomat pienet yllätyksetkin on saatu hoidettua loppuun. Theseus on suurin Kansalliskirjaston ylläpitämistä julkaisuarkistoista, eivätkä DSpaceen mahdollisesti sisältyvät bugit tule välttämättä esille testipalvelimilla tai muissa, Theseusta pienemmissä tuotantoympäristöissä. Näitä sekä joitakin suorituskykyyn liittyviä ongelmia korjailtiin heti päivityksen valmistuttua. Kansalliskirjaston Kiwiin on luvassa täydennettyä yleisohjeistusta julkaisuarkistoihin liittyen, vanha Theseus-virkailijan opas odottaa myös päivitystä.

Theseuksen etusivu on ehkä aavistuksen sinisempi kuin aiemmin vai olisiko peräti hieman harmaantunut? Graafinen taustakuva skaalautuu eri tavalla kuin aikaisemmin ja vaikutelma on rauhallisempi. Etusivun ulkoasun uudistuksen suunnittelua on tarkoitus alkaa työstää kunhan monet muut palveluun liittyvät keskeneräiset asiat on saatu ratkaistua. Kohta kunnioitettavan

10 vuoden iän saavuttava Theseus ei siis ole harmaantuva vanhus vaan dynaaminen kehittyvä palvelu, jonka responsiivinen käyttöliittymä skaalautuu käytettäväksi nyt myös mobiililaitteella.

Entäs SOLR?

Päivityksen myötä Theseuksen indeksointimekanismi vaihtui SOLR-pohjaiseksi. SOLR on avoimeen lähdekoodiin perustuva hakukonemoottori, joka toimii itsenäisenä palvelinohjelmistona. SOLR perustuu Apachen Lucene -hakumoottoriin ja sillä on itse asiassa jo aika pitkä historia itselläänkin, kehitystyö aloitettiin vuonna 2004. SOLRia käyttävät hakukoneenaan useat kävijämääriltään suuret ja sisällöltään laajat sivustot.

Indeksointimekanismin vaihtuminen toi päivitykseen haasteita, jotka Kansalliskirjaston mukaan ovat lopputuloksen arvoisia. Theseuksen sisältö ja yhtenäinen metadata saadaan paremmin näkyville käyttöliittymässä. SOLR ei kuitenkaan vaikuta ainoastaan indeksointiin, vaan myös Theseus-hakujen oletusoperaattori AND oli päivityksen yhteydessä vaihtunut OR-operaatööriksi, joka on SOLRin oletusoperaattori. Tämä saatiin kuitenkin jo palautettua takaisin alkuperäiseen AND-oletusoperaattoriin.

Näkyvimmit muutokset, hyvässä ja pahassa

Nostan esille muutamia kohtia, jotka Theseusta käyttäessään tulee varmasti huomaamaan. Alla olevassa kuvassa, jossa on uusi etusivulle tullut hakufasetti **Suuntautumisvaihtoehto**, on ensimmäisenä vaihtoehtona "III ". Jos tätä suuntautumisvaihtoehtoa klikkaa, ei saa hakutulokseksi mitään. Tämä on bugi, joka liittyy SOLR-hakumoottorin sisäiseen tallennusmuotoon eikä sitä ole helppo korjata. Voimme siis varautua siihen, että tämä "kolmen tolpan" suuntautumisvaihtoehto on näkyvissä em. hakufasetissa vielä pitkään.



Kuva 1. Suuntautumisvaihtoehto - hakufasetti

SOLR-hakumoottorin tekosiksi voidaan lukea myös selaushakujen hakutulostenäytöt, joissa nyt listataan ensimmäisiksi sellaiset tekijäkenttiin tallennetut tiedot, joissa on ensimmäisenä merkinä erikoismerkki. Aikaisempi hakumoottoriversio listasi erikoismerkit hakutuloksen loppuun. Tämän ominaisuuden avulla pystyy nyt helposti bongaamaan mahdolliset virhetallennukset, tekijälistaushan on todella pitkä...

Tekijöiden nimet
(toim.) [1]
(toim.), Stålbbyggnadsinstitutet [1]
, (Eds.) [1]
, Bela Gazdag [1]
, Iida Järvelä [1]
, Juha Laitila [1]
, Kohonen [1]
, Maija Grönroos [1]
, Mary Zambetakis [1]
, Niina Vänttinen [1]

Kuva 2. Tekijäselauksen hakutulostauksen alkupäätä

Kuva 2. Tekijäselauksen hakutulostauksen alkupäätä

Etusivun oikean reunan "Silmäile" –otsikon alta löytyvien uusien fasettien tekemät sisällön sorttaukset tulevat parhaiten esiin seuraavan esimerkin kautta. Jos klikkaat "Selaa Theseusta" –otsikon alta kohdasta Asiasanat, saat listauksen kaikista Theseukseen dc.keyword ja dc.subject –metadatakenttiin tallennetuista sanoista, siis opiskelijan tallentamat avainsanat ja kirjastohenkilökunnan tallentamat asiasanat aakkosjärjestyksessä, suluissa perässä näkyy sanan esiintymiskertojen määrä. Kuva havainnollistaa myös sen, että on merkitystä kirjoittaako sanan isolla vaiko pienellä alkukirjaimella:

Asiasana
a cappella [2]
a posteriori [2]
A(H1N1)-virus [4]
A* -algoritmi [1]
A-klinikat [12]
a-klinikat [1]
A-kliniker [1]
A-lehdet Oy [1]
A-Tuottajat [1]
A-vitamiini [1]

Kuva 3. Asiasanaselauksen alkupäätä

Asiasanojen sorttaus Silmäile-otsikon alla tuo ne näytölle lukumäärän mukaan, eli kuinka monta kertaa ko. asia/avainsanaa on Theseuksessa käytetty em. metadatakentissä. Yllätysmomentti

seuraakin, kun klikkaat kohdasta "...Katso lisää". Käyttäjän kuvitellessa, että sieltä näkee seuraavaksi eniten käytetyt asiasanat, saakin nähdäkseen aivan muuta:

Asiasana

markkinointi (3542)
kehittäminen (3538)
nuoret (2664)
laatu (2507)
lapset (2374)
suunnittelu (2275)
asiakastytyväisyys (2255)
kyselytutkimus (1986)
sosiaalinen media (1954)
hoitotyö (1881)
kvalitatiivinen tutkimus (1806)
työhyvinvointi (1772)
varhaiskasvatus (1762)
palvelut (1679)
yritykset (1642)
... Katso lisää

Viitteet 1-10 / 126730

"ei suositeltavat" -tuotteet (1)

"en meningsfull dag" (1)

"Finska Trädgårdsodlaren" (1)

"livsglädje" (1)

"Minun muistivihkoni" (1)

"New York tema" (1)

"Tavoittele terveyttä" -tuotos (1)

"tillräckligt bra" föräldraskap (1)

"värikylypy" (color bath) (1)

Moniammatillisuus # Vanhuspalvelut # Työhyvinvointi (1)

Kuva 4. Mitä tapahtuukaan, kun haluat katsoa lisää...

Tämä sorttaustuloksen näkymän muuttaminen on työn alla Kansalliskirjastossa, joten toivotaan että tämä on ohimenevä tilanne.

Loppu hyvin, kaikki hyvin

Nuo olivat päivityksen ikävämpiä yllätyksiä, mutta onneksi siihen sisältyy myös pelkästään positiivisia muutoksia. Erittäin toivottu uudistus etusivulla on parannettu tuki kansikuvien näytölle. PDF-tiedostoista generoidaan kansikuvat, jotka ainakin vielä toistaiseksi ilmestyvät pienellä viiveellä esiin. Tästä onkin tullut rukkaspostiin useita kysymyksiä, kun tallennuksen julkaisemisen jälkeen kansikuvan kohdalla näkyy vain harmaa laatikko, jossa lukee "No thumbnail". Kansalliskirjastossa ollaan tietoisia tästä viiveestä ja sitä on tarkoitus saada nopeammaksi myöhemmin.

Theseukseen tuli käyttöön myös hakukorostusominaisuus, joka käytännössä tarkoittaa sitä, että hakutuloksessa näytetään **lihavoituna** ne kohdat metadatatista (tiivistelmä) tai kokotekstistä, joihin haku osuu. Ominaisuuden ansiosta hakutuloksissa on nyt näkyvissä enemmän tietoa. Tämä helpottaa tiedonhakijaa, kun pystyy heti näkemään, onko hakutulos oikeassa asiayhteydessä. Hakuja pystyy myös rajaamaan uudella tavalla, kun ensimmäisen haun jälkeen saatua hakutulosta tarkentaa linkin "Näytä tarkemmat rajaukset" filttareiden avulla tai sorttaamalla ne haluamaansa järjestykseen esimerkiksi relevanssin, julkaisu- tai syöttöajankohdan tai nimekkeen mukaan.

Seuraava muutos Theseuksessa on toivon mukaan sen syöttölomakkeisiin liittyvä. Kansalliskirjaston metadatasuositus valmistui keväällä ja erityisesti rinnakkaistallennuksiin käytettävän lomake olisi sen pohjalta pikaisen päivityksen tarpeessa. Seuraava Julkaisuarkistojen asiantuntijatyöryhmän kokous on syyskuussa ja siellä kenties saadaan lisätietoa siitä kuinka julkaisuarkistojen metadatan yhtenäistämistyössä on tarkoitus ryhtyä etenemään. Tästä ja monesta muustakin Theseukseen liittyvästä asiasta tullaan kertomaan Kreodin tulevissa numeroissa, joihin Theseus-uutiskirje siirtyy Theseus-toimiston vakiopalstaksi.

Tiina Tolonen, informaatikko
Oulun ammattikorkeakoulu