

Apps4Finland-palkittu Eero Hyvönen: ”Datan julkaisumuotoon pitää kiinnittää jatkossa enemmän huomiota”

Näkökulmat Julkaistu 07.01.2015 klo 12:56 Kirjoittaja Päivi Rauste

Apps4Finland – avoimen datan innovaatiokilpailussa - avoimen tieteen ja tutkimuksen haastekategorian voitti tänä vuonna [Linked Data Finland - Kansallinen linkitetyn avoimen datan palvelualusta \(Linked Open Data\) työkaluineen ja sovelluksineen](#). Tavoitteena oli löytää työ, joka kuvastaa avoimen tieteen ja tutkimuksen periaatteiden toteutumista ja samalla tuo esille uudenlaista ajattelua ja toimintatapaa. Linked Data Finland -työryhmään kuuluivat: Eero Hyvönen, Jouni Tuominen ja Eetu Mäkelä.

Linked Data Finland on käynnistänyt ja edistänyt tutkimustietovarantojen hyödyntämistä, ja se toimii uranuurtajana semanttisten web-teknologioiden saralla. Hanke on osoittanut miten mahdollistetaan semanttisen tietosisällön automaattinen tuotanto, sisältöjen yhdistäminen, julkaiseminen käyttöön otettavina verkkopalveluina (webservice, mashup) sekä uusien kontekstiherkkien sovellusten tuottaminen.

Professori Eero Hyvönen, tuntuuko, että palvelun eteen tehty työ on tuottanut tulosta? ”Tätä Linked data -ajattelua on kehitetty jo pitkään. Linked Data Finland on FinnONTO-hankkeen jatkohanke, joka käynnistyi vuonna 2003. Kehitystyössä on ollut useita tahoja mukana. Viime aikoina monet asiat ovat menneet eteenpäin ja uusia tätä tekniikkaa hyödyntäviä sovelluksia on syntynyt. Mm. Digital Humanities -alue on ollut viime aikoina vahvasti kiinnostunut Linked data -tekniikasta. Samoin tieteellisen julkaisemisen Linked Science -idea on saanut vahvasti jalansijaa. Ideanahan on julkaista koneellisesti helposti käsiteltävää ja semanttisesti rikastettavissa olevaa, toisiin tietoverkkoihin yhdistettävää tietoa. Tämä sama idea on sovellettavissa kaikkialle, jossa tietoa käsitellään.”



Professori Eero Hyvönen

Avoimien datajoukkojen tuottaminen ja julkaiseminen

Linked data voi olla myös suljettua. Linked datan kehitys on ollut kuitenkin vahvasti sidottua tähän avoimien datajoukkojen tuottamiseen ja julkaisemiseen. ”Data on helpommin hyödynnettävissä, kun data julkaistaan tarkoituksenmukaisessa RDF-standardimuodossa, jonka jälkeen datan käyttö W3C-standardoitujen rajapintojen kuten SPARQL-palvelupisteen (endpoint) kautta on monesti paljon helpompaa”, tähdentää Hyvönen.

Hyvösen mukaan tiedon avaamishankkeissa ei ole kiinnitetty tarpeeksi huomiota siihen, missä muodossa data kannattaa avata. Enemmän voimavaroja kannattaisi keskittää hänen mukaansa

siihen, että jo avatut data, sekä tietysti tulevaisuudessa avattavat datat, olisivat mahdollisimman hyödynnettävässä ja tarkoituksenmukaisessa muodossa.

Tärkeimpiä asioita Linked Data Finland –projektissa on ollut yhteistyön fasilitointi. Organisaatiot tuottavat hyvin erilaista tietoa erilaisissa formaateissa. Tulevaisuudessa tarvitaan yhä enemmän yhteistyötä tiedon ja datan julkaisemisen alalla. Semattisen webin käyttö vaatii enemmän energiaa tiedon julkaisijoilta, koska dataa julkaistessa kannattaa miettiä, miten data kannattaa julkaista muiden ihmisten kannalta järkevästi. Datan julkaisumuodot tulevat yhä tärkeämpään rooliin jatkossa.

Seitsemän tähden linkitetyn datan palvelumalli

Linked Data Finland –työryhmä on kehittänyt Tim Barners-Leen luomaa linkitetyn datan viiden tähden mallia pidemmälle [seitsemän tähden linkitetyn datan palvelumalliksi](#). Tähdet on tarkoitettu kannustetähdiksi, joiden avulla ihmisten halutaan kiinnittävän huomiota asioihin, jotka ovat datan julkaisun kannalta arvokkaita. Linked Data -työryhmä jakaa tähtiä sen mukaan, miten hyvin määritellyt asiat on otettu huomioon. Työryhmän kehittämä kuudes tähti kuvaa sitä, miten hyvin datan skeemat (eli se, mitä data on) on julkaistu. Seitsemäs tähti myönnetään, mikäli data täyttää myös tähtiportaikon laatuvaatimukset.

Linked data -alustan hyödyt tutkijalle

Mitä etuja Linked Data Finland tarjoaa tutkijalle tieteellisten aineistojen käsittelyyn? ”Linked Data Finland tarjoaa alustan, jolla voidaan helposti julkaista tieteellisiä aineistoja. Aineistot voidaan linkittää muihin aineistoihin, jolloin syntyvien datasettien kautta pystytään vastaamaan sellaisiin kysymyksiin, joihin vastaaminen erillisten datasettien kautta olisi mahdotonta”, Eero Hyvönen selventää.

Hyvösen koneella on tällä hetkellä metatiedot Voltairen valistuksen aikana kirjoittamista 17 000:sta kirjeestä. Tämä liittyy Oxfordin sekä Stanfordin yliopistojen kanssa yhteistyössä tehtävään tutkimusprojektiin. Projektissa tutkitaan valistuksen ajan ajattelijoiden ja filosofien kirjeenvaihtoon liittyvää metadataa ja samalla sitä, miten ideoiden leviäminen tapahtui aikoina, jolloin postivaunut oli juuri keksitty ja tieteellinen ajattelu käynnistynyt. Linked Data Finlandin tausta-ajatuksena on humanistisen tutkimuksen tukeminen alustan kautta, jonka avulla datan yhdistäminen, analysoiminen ja visualisointi on mahdollista. Tutkitaan, voidaanko datasta löytää automaattisesti algoritmeilla jotakin sellaista, jota ihminen ei pystyisi jättimäisistä dataseteistä löytämään. Jos tieto vielä pystytään visualisoimaan järkevällä, kiinnostavalla ja ymmärrettävällä tavalla, tieto on paremmin levitettävissä ihmisten tietoisuuteen.

Linked Data Finland -palvelun tulevaisuus

Linked Data Finland -palvelua kehitetään jatkuvasti eteenpäin. Työryhmä kehittää uusia välineitä ja työkaluja sekä integroi niitä palveluun. Esimerkiksi tällä hetkellä on kehitteillä data-analytiikkavälineitä, automaattista datavalidointia sekä palvelu, jolla automaattista annotointia kehitetään ja näin aineiston tuottaminen tulee helpommaksi. Samalla otetaan käyttöön myös muiden kehittämää työkaluja ja välineitä. Lisäksi on kehitteillä online-kurssi datan avaajalle. Kurssi tarjoaa interaktiivisia harjoittelumahdollisuuksia sekä tukea datan avaamiseen ja linkittämiseen.

”Datan avaamisen kehittämisessä on tällä hetkellä rahoituksellinen musta aukko”, tuskailee Hyvönen. ”Julkinen sektori on Linked data -työtä tällä hetkellä pääasiassa kehittävä taho. Rahoituksen hakeminen esimerkiksi Tekesistä on vaikeaa, koska vielä ei ole kovin paljoa dataa ja palveluita, joita yritykset voisivat käyttää. Julkisen sektorin tuottama data pitäisi saada ensin auki ja käytettäväksi, jotta yritysmaailma voisi sitä hyödyntää. Toisaalta tutkimustyö on soveltavaa, eikä sovellu helposti Suomen Akatemian hankkeiksi. Suurin hyötyjä datan avaamisesta on tällä hetkellä Suomen talous kokonaisuutena, ei niinkään mikään yksittäinen yritys. Tämä on työtä maamme tietoinfrastruktuurin luomiseksi, joka sitten mahdollistaa kaikenlaista toimeliaisuutta vähän samaan tapaan kuin vaikkapa tieliikenneverkosto.” GPS-paikallissignaalin avaaminen on hyvä esimerkki datasta, joka avattiin vapaasti kaikkien käytettäväksi ja se on mahdollistanut valtavan määrän uusia sovelluksia ja uutta taloudellista toimintaa. Mihin kaikkeen Linked datan käyttö voikaan vielä luoda mahdollisuuksia tulevaisuudessa?

Lähteet:

Eero Hyvösen haastattelu 18.12.2014

<http://www.seco.tkk.fi/projects/ldf/>

<http://www.ldf.fi/>