

16.6.2014

Liikenne- ja viestintäministeriö
kirjaamo@lvm.fi, emil.asp@lvm.fi

Viite

Lausuntopyyntö LVM138/05/2014

Lausunto liikenne- ja viestintäministeriön raporttiluonnoksesta "Big data -strategia"

Aalto-yliopisto pitää varsin positiivisena aloitteena kansallisen big data -strategian kehittämistä. Big datan taustalla on maailmanlaajuinen nopeasti kehittyvä ilmiö, jossa datan tallennus- ja käsittelykapasiteetti on pitkään kasvanut eksponentiaalisesti hinnan samalla laskiessa, ja nyt ja lähitulevaisuudessa mahdollistaa aivan uudenlaisia palveluja. Datasta on tulossa strateginen resurssi ja kilpailuetu. On odotettavissa, että big data -teknologialla on erittäin vahva vaikutus yhteiskuntaan. On selvää, että Suomi ei voi jättäytyä jälkeen tästä kehityksestä. Kansainvälinen kilpailu tulee kuitenkin olemaan kovaa, eikä riitä, että meillä reagoidaan maailmalta tuleviin aloitteisiin. On erittäin tärkeää miettiä, mitkä ovat Suomen vahvuudet datan saatavuudessa, hallinnassa ja talletuksessa, analysoinnissa ja sovelluksissa.

Seuraavassa otetaan kantaa lausuntopyynnössä esitettyihin kolmeen kohtaan. Käytämme tässä edelleen termiä "big data", vaikka olisi tärkeää pyrkiä termin suomen-tamiseen; kotimaisten kielten tutkimuskeskus suosittelee käännöstä "massadata", jota pidämme onnistuneena.

Raportin erityiset toimenpide-ehdotukset

Aalto-yliopisto vastaa osaltaan big dataan liittyvästä tietotekniikan, tilastotieteen ja päätösanalyysin opetuksesta ja tutkimuksesta. Koulutuksen puolella big dataa on vaikea erottaa yleisemmästä data-analytiikan, ohjelmoinnin, rinnakkaisten järjestelmien, tiedonhallinnan, tilastotieteen ja päätösanalyysin opetuksesta. Big data on epämääräinen ja huonosti määritelty käsite; usein yritysten tarpeet saattavat liittyä sellaiseen datan käsittelyyn, jossa tietomassat eivät ole välttämättä erityisen suuria. Toisissa sovelluskohteissa datan suuri määrä taas vaatii aidosti uusia tapoja sen käsittelyyn, ja täten tarvitaan myös vahvaa big data -prosessointiympäristöjen sekä rinnakkaislaskennan osaamista perinteisen Suomessa erittäin vahvan data-analyysin osaamisen rinnalle. Big data -aineistot ovat usein myös muodoltaan uudenlaisia, esimerkiksi multimediadataa. Klassiset tilastolliset menetelmät eivät täten aina sel-laisinaan sovellu uudentyypisten aineistojen analysointiin, mikä edellyttää uusien tilastollisten ja laskennallisten menetelmien kehittämistä. Aalto-yliopisto kannattaa

16.6.2014

raportin näkemystä monialaisten big data -tiimien tärkeydestä, joissa matemaattinen, tilastotieteellinen ja päätösanalyttinen osaaminen yhdistyy ohjelmointi- ja algoritmivalmiuksiin, pilvilaskennan asiantuntemukseen, sovellusosaamiseen sekä tietosuoja- ja juridisten kysymysten hallintaan. Erityisesti teknologisten ja yhteiskuntatieteiden voimistuva uusi yhteistyö voi tuottaa merkittäviä innovaatioita.

Datan käytettävyys ja avoimuus on oleellisen tärkeä kysymys, koska big data -teknologia on datalähtöistä: ilman dataa ei voida saada tuloksia. On tärkeää korostaa julkisten tietovarantojen avoimen tiedon ohjelmaa ja sen pohjalle rakentuvaa uutta liiketoimintaa. Tutkimuksen ja opetuksen kannalta on tärkeää ymmärtää, että myös yritykset joutuvat ainakin jossakin määrin avaamaan tietovarastojaan, koska yritysten sisällä ei ole samaa asiantuntemusta data-analytiikasta kuin yliopistoissa. Menetelmäkehitykseen ja koulutuskäyttöön tarvitaan oikeaa konkreettista dataa. Suomen tekijänoikeuslaki antaa huonot mahdollisuudet tiedon louhintaan. Lakiin tulisi harkita rajoitussäännöstä, joka sitovasti mahdollistaisi tekijä- ja lähioikeuksien suojaaman materiaalin käytön ei-kaupalliseen tutkimukseen eli tutkimusrajoitussäännöstä nykyisen InfoSoc -direktiivin sallimissa rajoissa.

Monien suurien aineistojen kerääminen ja tallentaminen on tällä hetkellä liian satumanvaraista. Siksi on syytä kiinnittää enemmän huomiota siihen, mitä dataa on kerättävä ja millaisia päätöstehtäviä datan avulla pyritään tukemaan. Vaikka data onkin määritelmällisesti big data -tarkastelujen kulmakivi, sen kerääminen ei kuitenkaan ole tarkoitus sinänsä: se saa merkityksensä vasta silloin, kun dataa pystytään jalostamaan kehittyneillä laskennallisilla, tilastotieteellisillä ja päätösanalyttisillä menetelmillä niin, että näiden tarkastelujen pohjalta pystytään tekemään parempia päätöksiä.

Big data -tutkimusta tehdään Suomessa ja erityisesti Aalto-yliopistossa sekä Aalto-yliopiston ja Helsingin Yliopiston yhteisessä Helsinki Institute for Information Technology (HIIT) -tutkimuslaitoksessa, mutta tutkimuksen suhteellinen volyymi voisi olla suurempikin ajatellen alan kasvavaa merkitystä. Tervetulleena pidetään ehdotusta, että julkiset rahoittajat (Tekes, Suomen Akatemia) kohdistaisivat voimakkaammin ohjelmarahoitustaan big data -kehitykseen.

Raportti ehdottaa, että luodaan alan yrityksiä, tutkijoita ja asiantuntijoita kokoava avoin big data -klusteri. Aalto-yliopisto on tutkimuksen ja opetuksen johtavana toimijana valmis osallistumaan tähän. On tärkeää määritellä tällaisen klusterin toimintatapa, jotta vältetään päällekkäisyydet jo olemassa olevien yhteistoimintaverkostojen kanssa. Aalto suhtautuu myös positiivisesti kansainväliseen verkostoitumiseen, kunhan se tapahtuu kansallisista lähtökohdista alkaen.

Minkälaisiin big dataan liittyviin toimiin Aalto-yliopisto on ryhtynyt

Aallossa on v. 2009 suoritettuna tutkimusarvioinnin perusteella valittu neljä strategista tutkimuksen kohdealuetta, joista kahdessa, "Computation and modelling" sekä "ICT and media" on vahva kytkentä big data -tutkimukseen. Digitalisaatio on valittu Aallon fokusalueeksi, johon kuuluu suoraan 80 Aallon professoreista, ja toiset 80 sovellusten kautta. Hiljattain perustettu aloite "Aalto digi platform" koordinoi alan toimintoja, yhdessä HIIT:in ja EIT ICT Labsin kanssa.

Aallossa lähinnä tietotekniikan laitoksilla on jo pitkään opetettu erillisiä kursseja, jotka aiheeltaan liittyvät big dataan: tiedonhallinta, Web-tekniikat, rinnakkaiset järjestelmät ml. pilvipalvelut ja big data -prosessointiympäristöt, tiedon louhiminen ja data-analyysi. Samoin aiheeseen liittyvää tutkimustyötä on tehty pitkän aikaa mm. suurten teksti- ja multimediadokumenttikokoelmien hallinnan, bioinformatiikan, hajautettujen järjestelmien ja WWW-ontologioiden parissa. Aallon kauppakorkeakoululla on useita vuosia fokusoitu etenkin tekstinlouhintaan liittyvään sentimenttinetelmä- ja soveltavaan tutkimukseen. Syksystä 2014 alkaen on Aallon kaikissa maisteriohjelmassa mahdollista sisällyttää opintoihin monitieteinen sivuaine "Analytics

16.6.2014

and Data Science", joka vastaa suoraan big data -haasteisiin. Kauppakorkeakoulun puolella "Information and Service Management" -maisteriohjelman suuntautumisvaihtoehto "Business Analytics" on syksystä 2014 alkaen uudistettu vastaamaan big data -koulutustarpeisiin, sisältäen mm. uuden "Data Science for Business" -kurssin. Syksyllä 2015 alkavat Aallossa kokonaan uusitut maisteriohjelmat, joiden uudessa laajassa ICT-alan ohjelmassa "Tieto-, tietoliikenne- ja informaatiotekniikka" on tietotekniikan hakukohteessa eräänä erikoistumiskohteena "Big Data and Large-Scale Computing".

Aallossa toimivissa valtakunnallisissa tutkimuksen huippuyksiköissä on kaikissa vahva laskennallinen elementti. Erityisesti Laskennallisen päättelyn huippuyksikkö COIN (2012 – 2017) keskittyy vahvasti suurten data-aineistojen analyysi- ja päättelyongelmaan. Pääsovellukset ovat tiedonhaku ja bioinformatiikka. Myös vuoden 2013 lopussa päättynyt ALGODAN (Finnish Centre of Excellence for Algorithmic Data Analysis Research) -huippuyksikkö, joka jatkaa osittain nyt muulla rahoituksella, on keskittynyt juuri big data -aihepiiriin osalta keskeiseen algoritmisen data-analyysin aihepiiriin.

Minkälainen rooli voisi Aalto-yliopistolla olla kansallisessa big data -kehittämistyössä

Aalto-yliopisto on eittämättä Suomen vahvin opetus- ja tutkimusyksikkö tieto- ja viestintätekniologioiden alalla, ja tämän alan sisällä Aallossa ja sen edeltäjissä TKK:lla/HSE:llä on ollut pitkään vahvaa erikoistumista datalähtöiseen tietojenkäsittelyyn, jonka menetelmät ovat big data -analytiikan pohjana. Erityisesti Aalto-yliopisto, Helsingin yliopisto ja niiden yhteinen HIIT-tutkimuslaitos ovat monella indikaattorilla mitattuna maailmanlaajuisestikin kärkipaikalla koneoppimisen alalla, joka on eräs big data -alueen ydintekniikoita. Tätä vahvaa ydintä entisestään vahvistamalla voidaan luoda big data -alueelle Suomeen erityistä kilpailuetua. Aallolla on olemassa olevat vahvat verkostot alan yrityksiin ja muihin tutkimuslaitoksiin sekä vahvat kansainväliset yhteydet. Aalto-yliopisto on valmis osallistumaan mahdolliseen perustettavaan big data -klusteriin ja erityisesti kansalliseen data-analytiikan toimijoiden verkostoon, joka edistää big data -teknologiansiirtoa yrityksiin. Aalto-yliopisto kouluttaa big data -asiantuntijoita niin kandidaatti-, maisteri- kuin tohtoritasolla ja osaltaan vastaa data-analyysin, hajautetun laskennan, tilastomatematiikan ja päätösanalyysin asiantuntijoiden saatavuudesta suomalaisten yritysten palvelukseen. Aalto-yliopisto voi tarjota asiantuntemustaan kurssitarjontana, konsultointiin, kansallisen big data -infrastruktuurin pystyttämiseen ja standardien kehittämiseen. Standardien kehittämisessä mm. Aalto-yliopiston Real Time Economy -osaamiskeskuksella on arvokasta kokemusta liittyen sähköisten tilinpäätöstietojen XBRL-standardisointiin.

Ilkka Niemelä
provosti
Aalto-yliopisto