

Humanistit datan kimpussa: putsaamista ja pöyhintää koko viikko

Fakta Julkaistu 26.05.2016 klo 09:47 Kirjoittaja Jessica Parland-von Essen

Helsingin yliopistossa järjestettiin äskettäin toisen kerran digitaalisten ihmistieteiden käytännönläheinen kurssi Helsinki Digital Humanities Hackathon. Viikon ajan neljä poikkitieteellistä tiimiä pakersi sähköisten aineistojen ja menetelmien sekä humanististen ja yhteiskuntatieteellisten kysymysten parissa. Tuloksena syntyi näyttäviä visualisointeja, uusia kysymyksiä ja oivalluksia sekä mukavasti uutta avointa koodia. Avoimuus ja yhteistyö ovatkin tällä uudella tutkimusalueella itsestäänselviä lähtökohtia, joiden avulla kehitys pysyy nopeana. Uutta tänä vuonna oli visuaalisten aineistojen laajempi käyttö sekä ihmisen tukema koneoppiminen, jolla hiottiin useita menetelmäsovelluksia. Paljon liittyy digitaaliseen ihmistieteeseen aina myös aineiston analysointia, lähilukua ja siistimistä, mikä tänäkin vuonna teetti runsaasti töitä.



Kaupunkihistoriaa tutkittiin muun muassa käyttämällä Finnan avointa rajapintaa. Länsi-Pasilan historia tarjoaakin todellista dramatiikkaa ja mielenkiintoisia kysymyksiä.

Ilahduttavaa oli laaja osaanottajapiiri. Tänä vuonna mukana oli myös Yle, jonka ohjelmien metatietoja sekä pöyhittiin että rikastettiin huippututkijoiden toimesta. Kotimaiset aineistojen tarjoajat, Kansalliskirjasto ja Yle, olivatkin kurssilaisten tukena ja oppivat samalla lisää tutkijoiden tarpeista ja omista aineistoistaan. Näin avoimuus voi olla hyvinkin palkitsevaa puolin ja toisin. Nyt esimerkiksi Yle sai arkistoonsa asiantuntevien mediatutkijoiden tuottamia uusia metatietoja.

Englantilaista korpusta työstettäessä Aalto-yliopiston Jukka Suomela kehitti hieman brutaalilta tuntuvan, mutta mielenkiintoisen normalisointimenetelmän, jonka avulla voidaan ohittaa osa vanhojen tekstien epävakasta kirjoitusasusta johtuvista ongelmista.

Kaikkien ryhmien töiden esittelyt ja linkit Githubiin löytyvät [tapahtuman sivun](#) lopusta.