

Petri Paju

Tarjolla digitaalisia työvälineitä humanisteille

Digitaalisen humanistisen tutkimuksen pohjoismainen DHN-konferenssi (*Digital Humanities in the Nordic Countries*) järjestettiin Göteborgissa maaliskuun puolivälissä 2017. Tapaaminen oli järjestyksessä toinen lajiaan. Sarjan ensimmäinen konferenssi oli pidetty vuonna 2016 Oslossa. Kumpikin kokous on tarjonnut varsin edustavan läpileikkauksen digi-humanistien ja aiheeseen liittyvien instituutioiden kuten kansalliskirjastojen hankkeisiin ja kiinnostuksenkohteisiin Pohjoismaissa. Mukana on ollut laaja kirjo niin kielten ja historian tutkijoita kuin sovellusten kehittäjiä sekä arkistojen ja kirjastojen edustajia. Göteborgissa nähtiin joitakin esiintyjä myös niin Baltiasta kuin kauempaa Britanniasta, Saksasta ja Venäjältä. Seuraavaksi konferenssi saadaan Helsinkiin.

Pääpuhujia oli neljä. Ensin esiintyivät kirjallisuudentutkijat Peter Leonard (Yale) ja Tim Tangherlini (UCLA) Yhdysvalloista. He puhuivat ”makroskoopin” (*developing a “macroscope”*) kehittämisestä ja soveltamisesta Pohjoismaiden kirjallisuuksien tarkasteluun. Kaksikko hahmotti eri tapoja käyttää ja navigoida suuria tekstikokonaisuuksia *distant reading* -menetelmän keinoin ja näytti kuinka ohjelmistotyökalujen avulla oli ollut mahdollista löytää yllättäviä, kauan siten unohduneita kirjallisia yhteyksiä ja yhtäläisyyksiä valtavasta digitoidusta tekstimassasta. Dolly Jørgensen (Uumajan yliopisto) puhui siitä, kuinka digitaaliset keinot ovat muuttamassa tapojamme nähdä, tietää ja täydentää (*augment*, kuten *augmented reality* -sovellukset) luontoa, ja kehotti tutkimaan digitaalisuutta tästäkin näkökulmasta. Katja Kwastek (Vrije University Amsterdam) esitti katsauksen keskusteluihin taidehistoriasta, taitelijoihin, visuaalisuudesta ja digitaalisuudesta.

Historiantutkijoiden kannalta kiinnostavia uutisia saatiin digitaalisista sanomalehtikokeelmista eri Pohjoismaissa ja niiden automaattisesta nimentunnistuksesta Suomessa. Kimmo Ketunen Kansalliskirjastosta kertoi, että esikuvana on italialaisen *La Stampa*-sanomalehden NER-toiminnot (*Named Entity Recognition*), joihin voi tutustua verkossa.¹ Sen vasen sivupalkki kertoo esimerkiksi, mitä kaupunkeja tai maita jollakin haulla tuloksina saaduissa artikkeleissa on mainittu. Suomalaisella sanomalehtiaineistolla tehdyt NER-kokeet vaativat vielä kehittämistä tuottaakseen riittävän luotettavia tuloksia. Tämä johtuu erityisesti aineiston OCR-(epä)tarkkuudesta, joka vaikeuttaa nimien tunnistamista. Sen sijaan EU-hanke READ, jossa kehitetään vanhoja käsinkirjoitettuja tekstejä koneellisesti tunnistavaa, oppivaa ohjelmaa, on tuottanut erinomaisen lupaavia alustavia tuloksia. Hyvien tulosten edellytyksenä on, että analyysin pohjana on riittävästi tarkastettua aineistoa, jonka avulla opettaa ohjelmalle tietty käsiä.² Maria Kallio Kansallisarkistosta esitteli osaltaan näitä uutisia konferenssia välittömästi edeltäneessä työpajassa.

Yleisemmän katsauksen suomalaisen digitaaliseen historiantutkimukseen tarjosi yhdessä professori Mats Fridlundin kanssa pitämäni esitelmä, joka perustui Koneen Säätiön rahoittamaan hankkeeseen kartoittaa digitaalisen historiantutkimuksen kehitystä Suomessa.³ Kerroimme myös hankkeen jatkosta, jossa syvennytään esimerkiksi digitaalisen humanistisen tutkimuksen laboratorioden toimintaan. Suomessa tätä kehitystyötä vie eteenpäin erityisesti Helsingin yliopiston HELDIG, Helsinki Centre for Digital Humanities, joka perustettiin vuonna 2016.

Suurimman vaikutuksen tekivät kokeet, joita oli tehty Ruotsissa Tukholman Tekniikan museon tilauksesta. Professori Pelle Snickarsin (Uumajan yliopisto) johdolla oli digitoitu erilaisia aineistoja jatko-työskentelyn pohjaksi. Näistä varmasti futuristisin kokeilu oli kolmiulotteisesti mallinnetut

1. Lehden sivut, <http://www.archiviolastampa.it/> (25.8.2017).

2. Ohjelman alustava versio on jo saatavilla, <https://transkribus.eu/Transkribus/> (25.8.2017).

3. Ks. <https://digihistfinlandroadmapblog.wordpress.com/> (25.8.2017).

Dosentti **Petri Paju** työskentelee projektitutkijana Aalto-yliopistossa. Sähköposti: petri.paju@aalto.fi.

puiset, mekaaniset mallit 1700-luvulta. Ne muodostivat Christopher Polhemin, ”ruotsalaisen tekniikan isän”, *Laboratorium mechanicumin* ytimen. Malleilla oli ja on tarkoitus opetella erilaisen liikkeen aikaansaamista esimerkiksi hammaspyörästöjen avulla. Ammattimainen animaattori oli taituroinut Polhemin ”mekaanisista aakkosista” näyttävät 3D-mallit. Niiden käyttöä oli projektissa kokeiltu virtuaaliodellisuudessa, jossa 3D-laite saattoi (yllättäen?) myös rikkoutua eli vaikkapa pudottaa jonkin liikkuvan osansa. Sama *Digitala Modeller*-projekti on myös digitoinut Tekniikan museon *Dædalus*-vuosikirjat (*Tekniska museets årsbok*) vuodesta 1931 vuoteen 2015 verkkoon kaikkien saataville.⁴ Julkaisun sisältö kiinnostaa ainakin tekniikan ja teollisuuden historiaa tutkivia.

Kokonaisuudessaan esitysten aihekirjoa voi sanoa laajaksi. Jatko-opiskelija ja egyptologi Kölnistä (Lena Tambs) esimerkiksi kertoi tutkimuksestaan, jossa hän hahmottaa tietyn sotilasleirin elämää ja erityisesti kaksikielisyyden kehitystä tutkimalla paikalta kaivettuja teksti- ja arkistolöydöksiä verkostoanalyysin työkaluin. Venäläisten osallistujien posterit esitteli Prozhito-projektin ja sivuston, johon tutkimusryhmä kerää yksityisten päiväkirjoja verkkoarkistoon kaikkien luettaviksi ja avaa näin arkielämän näkökulmia Venäjän historiaan.

Pohjoismainen paneeli www-sivujen arkistoinnista suorastaan pakotti ajattelemaan sitä, millaisen käsityksen eri maiden kansallisarkistojen aineistot mahtavat tulevaisuudessa antaa nykypäivän verkkomaailmasta. Ainakaan toistaiseksi kuva ei liene kovin edustava, niin monia kysymyksiä ja hidasteita tuli esille alkaen lainsäädännöstä ja aineiston keräämisestä sen esittämiseen. Automatisoitua sekä ohjattua keruuta tehdään joka maassa lähinnä kansallisten verkkotunnusten (kuten .fi Suomessa, .dk Tanskassa)

rajoissa. Nykyiset kokoelmat ovat selattavissa (jos ovat) tietyillä tietokoneilla arkistojen tiloissa, mikä on sinänsä mahtava paradoksi suhteessa internetin kaikkialle ulottuvuuteen. Toisin kuin aiemmin arveltiin ja joskus edelleen esitetään, tietoverkot eivät säilö kaikkea, vaan osalla verkkosivuista ja -aineistoista on hyvinkin lyhyt elinkaari. Yksi johtopäätös esitelmistä olikin se, että omalle tutkimukselle erityisen tärkeät verkkosällöt kannattaa pyrkiä arkistomaan itse, sillä muuten ne voivat jokin päivä olla saavuttamattomissa.

Monipuolisesta konferenssista ei välttämättä syntynyt selkeää yleiskuvaa, mutta sen verkkosivujen abstraktit antavat lisätietoa runsaasta tarjonnasta ja ajankohtaisista kysymyksistä. Kuulun perusteella on selvää, että Pohjoismaiden muisti-instituutioilla ja tutkijoilla on jatkuvasti opittavaa toistensa ratkaisuihin ja on eduksi, että arkisto- ja kirjasto-alan asiantuntijat sekä (muut) tutkijat kokoontuvat vaihtamaan ajatuksia kokoelmiensa ja alojensa tulevaisuutta muovaavista digitaalisista järjestelmistä. Kaiken kaikkiaan on perusteltua odottaa suurella mielenkiinnolla seuraavan pohjoismaisen digitaalisen humanistisen tutkimuksen konferenssia Helsingissä 7–9. maaliskuuta 2018. Yksi konferenssikutsun teemoista on historiantutkimus.

Lähteet ja linkit

Yllä mainitut projektit ja lehtiarkisto:

<http://www.archiviolastampa.it/>
<https://digihiistfinlandroadmapblog.wordpress.com/>
<http://digitalamodeller.se/>
<http://prozhito.org/>
<https://transkribus.eu/Transkribus/>

Vuoden 2017 DHN-konferenssin abstraktit:

<http://dhn2017.eu/abstracts/>

4. Hankkeen verkkosivut, <http://digitalamodeller.se/> (25.8.2017).