

SUKUKIELTEN DIGITOINTIPROJEKTI JA KIRJALLISUUSPANKKI ESIMERKKEINÄ TUTKIJAYHTEISTYÖSTÄ

Jussi-Pekka Hakkarainen

Kirjaston tori

Porthania, P674

13.3.2013

SUKUKIELTEN DIGITOINTIPROJEKTI

- Pilottikausi käyntiin heinäkuussa 2012, loppuu lokakuussa 2013.
- Hankkeen aikana Kansalliskirjasto tuottaa Koneen Säätiön Kieliohjelmassa hyödynnettävää tutkimusaineistoa ja –infrastruktuurin.
- Tutkimuskirjasto vastaa yhteistyöstä niin kotimaisten kuin ulkomaistenkin yhteistyötahojen kanssa.
- Digitoidaan suomalais-ugrilaisia kieliaineistoja (inkeroinen, vepsä, mordvalais- ja marilaiskielet) Venäjän Kansalliskirjaston kokoelmista, yhteensä 128 monografiaa (17 000 sivua), oppikirjoja ja sanakirjoja sekä lähes 20 000 sanomalehtisivua 1920- ja 1930-luvuilta.

SUKUKIELTEN DIGITOINTIPROJEKTI

- Sopimusneuvottelut venäläisten partnerien (Venäjän Kansalliskirjasto, Pietari; National Library Resource, Moskova) kanssa loka-marraskuussa.
- Tekijänoikeuksien selvitystyö käyntiin joulukuussa, kesto 90 työpäivää, aineiston luovutus selvitystyön jälkeen.
- Tekijänoikeudet kansalliskirjastoille mahdollisimman laajana:
 - 1) oikeuden valmistaa kappaleita tästä teoksesta.
 - 2) oikeuden asettaa tämän teoksen kappaleita yleisön saataville
 - 3) oikeuden myöntää kolmansille osapuolille lupa valmistaa kappaleita tästä teoksesta ja asettaa ne yleisön saataville.
- Aineisto julkaistaan projektin omassa Dspace-instanssissa.

KIRJALLISUUSPANKKI

- Taustat vuodelta 2009, käyntiin elokuussa 2012
- Kansalliskirjasto, Helsingin yliopiston kirjasto sekä Helsingin yliopiston Suomen kielen, suomalais-ugrialaisten ja pohjoismaisten kielten ja kirjallisuuksien laitos (Kotimainen kirjallisuus)
- Neljä rahoittajatahoa: OKM, Koneen Säätiö, Kulttuuriperintörahaston hoitokunta ja Helsingin yliopisto
- Hankkeen materiaalina ovat vuosina 2008–2009 digitoidut 600 kaunokirjalliset teokset – jo Doriassa.

KIRJALLISUUSPANKKI

- jakautuu neljään toiminnalliseen ulottuvuuteen, kenttään, jotka palvelevat erilaisia tiedontarpeita:
 - 1) Klassikkokirjasto PDF- ja OCR-muodossa (yleisö/tutkimusinfra)
 - 2) Bibliografinen tiedosto (tutkimusinfra)
 - 3) Kirjailijatiedosto (yleisöinfra)
 - 4) Kotimaisen kirjallisuuden digitaalinen vuosikirja (tutkimusinfra)
- Pilotin keskeinen tavoite on saattaa Kansalliskirjaston aiemmin tuottama digitaalinen aineisto sellaiseen muotoon, että sitä voidaan muokata ja siihen voidaan liittää metadataa.

OCR-EDITORI

- Kansalliskirjasto kehittää kielentutkijoiden käyttöön tarkoitettun työkalun, jonka avulla digitoidusta aineistosta automaattisesti tunnistettua tekstiä voidaan korjailla.
- OCR-editorin kehittäminen aloitettu KVP:ssa.
- Rahoitus sekä Sukukielten digitointiprojektista että Kirjallisuuspankista.
- Automaattisen tunnistuksen tulokset ovat puutteellisia kielten puutteellisten korpusten ja painojäljen huonon laadun takia.
- Kehitystyö yhteistyössä Kirjallisuuspankki-hankkeen ja suomalais-ugrilaisten kielten tutkijoiden kanssa.
- Tavoitteena on käytännössä virheetön tekstikorpus.

YHTEISHAKEMUKSET – LYHENEKÖ MATKA TUTKIJAN TYÖPÖYDÄLLE?

- Yhteydenotto Turun yliopiston Volgan alueen kielten tutkimusyksikköön Sukukielten digitointiprojektin kautta syksyllä 2012.
- Alustava kehys mahdollisista yhteistyömuodoista viime jouluna.
- Helmikuu 2013: tutustutaan organisaatioon, tutkijoiden työskentelykäytänteisiin ja paikallisiin resursseihin ja päätetään yhdessä mitä haetaan ja mihin haettava raha kohdennetaan.
- Apurahahakemuksen valmistelu yhdessä tutkijoiden kanssa.
- Aikatauluttaminen keskeisellä sijalla: hankkeen koordinointi vie tutkijoiden aikaa.

YHTEISHAKEMUKSET – LYHENEKÖ MATKA TUTKIJAN TYÖPÖYDÄLLE?

- Mietittiin yhdessä millaiseen tutkimukselliseen kenttään hakemus ja mahdollinen osahanke asettuu sekä lyhyellä että pidemmällä aikavälillä.
- Mutta toisaalta suunniteltiin yhdessä, miten osahanke voisi laajentua varsinaiseen jatkohankkeeseen vuosille 2014-2016.
- Humanistisen tutkimuksen rahoituksen luonne ja sen huomioon ottaminen.
- Ei sisällä esim. post doc- tai väitöskirjahankkeita, joten rahoituksen tulee tukea nykyisen tutkijakunnan intressejä, eli tutkimusteemat eivät mene uusiksi.
- Pienet purot ja yhteistyö välttämättömiä tutkimuksentekijöille: ”miljoonarahoitukset ovat utopiaa tällä tutkimuksen kentällä.”

LISÄTIETOJA

Jussi-Pekka Hakkarainen

jussi-pekka.hakkarainen@helsinki.fi

050 363 9223

Projektin verkkosivu siniseltä linjalta:

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

[kokoelmatjapalvelut/digitaalisetkokoelmat/finnougric.html](http://www.kansalliskirjasto.fi/kokoelmatjapalvelut/digitaalisetkokoelmat/finnougric.html)