

Grammatikka I oikiakirjutos.

oisesse klassaasas annettuun sanomustilittelu kertaamiin. Uta sanomuksen soderzanja ja sanoinn yhtlehytti sano- si. Upraznenjoja sanomuksiin laadimisces eri teemoille kochittämistöön yhtlehes.

rostoi sanomus, lauhitu sannoist, kummat vastajaat kysy- e: **ken? mikä?** (subjekta) i sannoist, mit vastajaat kysy- e: **ken tekköö?** mitä tekköö? (predikaattia). Subjektian kaagan keskinäiset ottoesjenat.

rajennettu sanomus. Eroittammiin sanomuksest sanoja, alittaa subjektian i predikaattia.

es klassaasas mäntylin predmettasannoin kertaamiin. nmiho nimisanan kera, ku.npa osoittaa predmettoja i lamliisellä, javienjoja, tapahtummiin).

misannoin muuttuamiin sanomuksees. (Kala ono mukuisa kalan pyytö tapalittuu ympär vootta. Keskiakesäl vällä ää kallaa j. n. e.) Sanan pohjan ja sanan lopua eroit- sanast.

misena subjektian i toimitussananna. Nimsisanan lopun nmiin, kons se on subjektian ja kons se ono toimitus- jo.

isikkö ja monikko. Nominetiva, ku forma, kons möö mä predmettaa. Toiset padezat. Padezoiin niinet. Pade- zoiin.

Tuustutamiin teesannoin kera. Tekosanan ponnates tees- mukkaa. Tekosannoin aikaformat tapalittuva aika (presens tapalittuat aika (imperfekta), sloonoi tuleva aika (noisen teke- mää). Tekosanan infinitiva format: a) 1-n infinitiva mitä tehdä, kytää, pyytää b) 2-in infinitiva, mitä tekemää? tekemä kytäämä, pyytäämä. Tekosanan loput monikoo i yksiköös.

11. Yhissetyt sanat, niijen oikiakirjutos i lauttuun.

NELJÄS KLAASSA.

Lukemiin.

1. Präntätyin ja käsikirjutoslekstan äänee lukemiin (tänne voojen töön jatkuamiin vähä rokkemman materlalanka). Tuntu merkiiä luetust materialast meellin, ajatuksiin eri o.tenkej. Oppiissa lukemaa hyväst äänee hudozestvennoita kirjaa. Tuntu valmistus itseksee lukemaa äänne väelle.

2. Toittumus tööää kirjanka. Lukemiin itseksee opeitaja annettua tekstaa täll itseen valittua tekstaa. Tuntu tehi mu sloppanoja luetust tekstast.

3. Kirjaa juko osii ja glovoihe. Tuntu antaa omija zag- violta luetule tekstalle. Tuntu jakkaa osiitsee luettu tekstaa antaa omat zaglaviat eri osille. Plaanaa laittimiin.

Tuntu löyää itäkäämykkest glavnoit paikat ja vtorostepe- nit paikat. Tuntu situu neel kostonää. Tuntu niijen keskiäsi-

SUOMALAIS-UGRILAISTEN AINEISTOJEN DIGITOINNIN PILOTTIPROJEKTI (2012–2013)

TIIVISTELMÄ

Tässä esitteessä esitetään tuotantovälineet ja -prosessi sekä organisaatio suomalais-ugrialaisten aineistojen digitoinnin pilottihankkeelle vuosille 2012–2013. Pilottihanke on osa Koneen Säätiössä valmisteltua kieliohjelman hankekokoaisuutta ja siihen viitataan Koneen Säätiön Kieliohjelmassa 2012–2016 otsikolla Pietarin aineistojen digitoinnin pilottihanke (2012–2013), tästedes tässä yhteydessä vain pilottihanke.

Pilottihanketta varten digitoitava aineisto julkaistaan Kansalliskirjaston ylläpitämässä ja avoimeen lähdekoodiin pohjautuvassa julkaisujärjestelmässä, jossa esitettävä digitaalinen aineisto saatetaan avoimeen kansalaiskäyttöön sekä taataan tiedeyhteisölle mahdollisuus näiden kieliaineistojen hyödyntämiseen tutkimuksessa. Lisäksi aineiston käyttöön asettamisen yhteydessä luodaan tarkoituksenmukaiset välineet ja edellytykset talkoistamispalveluille.

Selvityksessä ehdotetaan työnjaoksi, että pilottihankkeen toteuttamisesta ja koordinoinnista vastaa erikseen nimitettävä projektipäällikkö, joka sijoittuu Kansalliskirjastoon ja raportoi toiminnasta suoraan pilottihankkeen ohjausryhmälle. Projektipäällikkö vastaa Venäjän Kansalliskirjastossa (Pietari) tehtävän digitoinnin laatumäärittysten noudattamisesta. Hän valvoo aineiston käyttöön asettamista Kansalliskirjaston julkaisujärjestelmässä ja aineistoa koskevien tekijänoikeussäädösten noudattamista sekä koordinoi Kansalliskirjastossa suoritettavia työvaiheita. Lisäksi pilottihankkeen viestintä kuuluu projektipäällikön työtehtäviin yhteistyössä Kansalliskirjaston viestintäyksikön kanssa. Julkaisujärjestelmän ylläpidosta ja teknisestä kehittämisestä huolehtii Kansalliskirjaston Kirjastoverkkopalvelut ja kuvailu- ja formaattityöstä Kansalliskirjaston Tutkimuskirjastopalvelut. Kaikki projektista syntyvät kustannukset, mukaan lukien palkkakulut, katetaan Koneen Säätiöltä anottavalla apurahalla. Kansalliskirjasto huolehtii työntekijöiden työvälineistä, työpisteistä ja muusta välittömästä infrastruktuurista.

Selvityksen perusteella Koneen säätiöltä on otettiin 356 040 euron suuruinen apuraha pilottihankkeen toteuttamiseksi aikavälillä heinäkuu 2012–joulukuu 2013. Pilottihankkeen avulla kehitetään toimintatapaa ja teknistä toteutusta, joita on tarkoitus hyödyntää muiden aineistojen suhteen jatkohankkeessa vuosina 2013–2016.

OSAPUOLET JA ORGANISAATIO

Kansalliskirjasto tuottaa Koneen Säätiön kieliohjelmassa tarvittavan tutkimusaineiston ja infrastruktuurin Koneen Säätiöltä haettavalla apurahalla. Kansalliskirjasto vastaa myös yhteistyöstä Venäjän Kansalliskirjaston (Pietari) ja Helsingin yliopiston kirjaston kanssa. Kansainvälistä työnjakoa ohjaa Kansalliskirjaston ja Venäjän Kansalliskirjaston välinen yhteistyösopimus, jonka perusteella pilottihanketta varten neuvotellaan erikseen keväällä 2012 kirjas-tojen välinen sopimus, missä määritellään hankkeen puitteissa toteutettavat kummankin osapuolen vastuut ja velvollisuudet ehtoineen. Aineistojen tekijänoikeuksien selvittämisestä ja korvaamisesta sovitaan erikseen Kansalliskirjaston ja venäläisen kirjastoalan yhteisen tekijänoikeuksia selvittävän toimiston National Library Resourcen (NLR, venäjäksi Национальный библиотечный ресурс, НБР) välillä.

Pilottihanketta varten luodaan ohjausryhmä. Ohjausryhmä koostuu kotimaisten osapuolten ja tieteen tekijöiden edustajista siten, että myös kirjastoalan erityisosaaminen on ohjausryhmässä edustettuna. Kansalliskirjasto apurahan vastaanottajana vastaa pilottihankkeen raportoinnista ja tulosten esittelemisestä Koneen Säätiölle.

Ohjausryhmän on tarkoitus tukea hanketta varten palkattavaa projektipäällikköä, joka raportoi ohjausryhmälle projektin edistymisestä. Projektipäällikön sijoittuminen Kansalliskirjaston organisaation sisälle antaa hänelle parhaat mahdollisuudet toteuttaa pilottihanketta sekä valvoa hankkeen työvaiheita ja edistymistä. Hankkeen alkuvaiheessa, alkaen heinäkuusta 2012 projektipäällikkö vastaa pilottihankkeen yleisestä suunnittelusta, budjetoinnista, sopimusten solmimisesta hankkeen eri osapuolien ja muiden tarvittavien tahojen kanssa sekä alkuvaiheen tietojärjestelmäkoulutuksesta ja yhteistyökäytänteiden sopimisesta niin kotimaisten kuin ulkomaisten osapuolien kanssa.

Möhempiin projektipäällikön tehtäviin kuuluvat digitoitavan aineiston laadunvalvonta, työprosessien suunnittelu, käyttöön asettamisen valvonta, talkoistamispalveluiden käyttäjähallinta sekä tekijänoikeuksien valvonta. Projektipäällikkö vastaa myös yhteistyöstä tieteen tekijöiden kanssa. Projektin viestintä ja sen järjestäminen yhteistyössä Kansalliskirjaston viestinnän kanssa kuuluvat niin ikään projektipäällikön tehtäviin. Laskutuksiin ja rahaliikenteeseen liittyen projektipäällikkö työskentelee yhteistyössä Kansalliskirjaston taloushallinnon kanssa.

Projektipäällikön vastuiden ja työtehtävien lisäksi hankkeen eri vaiheissa tarvitaan erilaista kirjastoalan teknistä osaamista, jollaisen osaamisen avulla julkaisujärjestelmien ylläpitoa ja kehittämistä voidaan toteuttaa. Lisäksi pilottihankkeessa tarvitaan kirjastoalan kuvailu- ja metadatakokemusta, erityisesti kohdistuen hankkeen poikkeaviin metadatakäytänteisiin ja formaatteihin.

AINEISTO

Koneen säätöön edustajat ovat syksyllä 2011 määritelleet pilottihankkeessa digitoitavan suomalais-ugrilaisen aineiston, joka kuuluu Venäjän kansalliskirjaston (Pietari) kokoelmiin. Tämä tutkijoiden valitsema suomalais-ugrilaisilla kielillä julkaistu aineisto digitoidaan Pietarissa Venäjän Kansalliskirjastossa ja toimitetaan Kansalliskirjastolle digitaalisessa muodossa. Valikoitu aineisto sisältää noin 7000 sivua inkeroisten- ja vepsänkielisiä julkaisuja, yhteensä 66 monografiaa, jotka ovat pääsääntöisesti oppikirjoja ja sanakirjoja. Monografioiden lisäksi digitoidaan lähes 20 000 sivua marilaisia ja mordvalaisia sanomalehtiä, jotka on julkaistu pääsääntöisesti 1920- ja 1930-luvuilla.

Pilottihankkeen aikana digitoitavan aineiston teknisiä vaatimuksia on määritelty yhdessä Venäjän Kansalliskirjaston kanssa. Koska aineistoja ei voida luovuttaa Suomeen digitoitavaksi, digitointi suoritetaan Venäjän Kansalliskirjastossa ja heidän digitoimansa aineistot toimitetaan edelleen Kansalliskirjastolle. Monografiat digitoidaan harmaasävyskannattuna 300 dpi tarkkuuteen, kun taas sanomalehtien resoluutioksi on haluttu määritellä 400 dpi. Primääriskannaus suoritetaan TIFF-kuvaformaattiin ja skannattujen kuvien yhteyteen lisätään automaattisen tekstintunnistamiseen (OCR, optical character recognition) vaadittavat elementit. OCR-digitoitu teksti liitetään kuvatiedostoon ja nämä elementit liitetään yhdeksi tiedostoksi, PDF-formaattiin, joka on myös aineiston esitysformaatti julkaisujärjestelmässä. Mahdollista myöhempää käsittelyä varten myös TIFF-muotoiset kuvat luovutetaan Kansalliskirjastolle.

Venäjän Kansalliskirjasto toimittaa myös digitoitavien aineistojen bibliografiset tiedot Kansalliskirjastossa (julkaisujärjestelmässä) tapahtuvaa aineistonkuvailua varten. Bibliografisia tietueita ei sellaisenaan voida hyödyntää julkaisujärjestelmässä, sillä venäläiset kirjastot käyttävät aineistojen kuvailuformaattina RusMarcia, kun taas pilottihankkeessa käytettävä digitaalisten aineistojen kuvailuformaatti on DublinCore, joka on digitaalisten julkaisujen kuvailuun kehitetty metadataformaatti. Luetteloinnin yhteydessä bibliografista metadataa

muokataan vastaamaan DublinCoren kuvailumääritelmiä digitaalisesta aineistosta. Pilottihankkeessa digitoitavaa aineistoa ei luetteloida paikallistietokantoihin, vaan aineiston saavuttavuutta edistetään mahdollistamalla sen löydettävyys Kansallisen Digitaalisen Kirjaston (KDK) ja mahdollisesti Europeanan käyttöliittymissä, niiltä osin kuin se on tekijänoikeuksien puitteissa mahdollista.

JÄRJESTELMÄT JA RAJAPINNAT

Pilottihankkeessa digitoitu aineisto julkaistaan Kansalliskirjaston ylläpitämässä julkaisujärjestelmässä, joka perustuu avoimen lähdekoodin DSpace-ohjelmistoon. Pilottihanketta varten perustetaan muista Kansalliskirjaston digitaalisista kokoelmista erillinen DSpace-instanssi (julkaisujärjestelmäympäristö), jonka avulla voidaan vastata niin tutkijakunnan kuin kansalaistieteen tarpeisiin. Rääätälöidyn DSpace-instanssin myötä kehitetään kansalaistieteen työkaluja sekä luoda paremmat edellytykset hanketta varten kehitettävälle toiminnallisuksille. Eräs pilottihankkeen aikana toteutettava kansalaistieteen työkalu on ocr-tekstieditori, jonka avulla julkaistun aineiston ocr-tekstiä voidaan korjata ja muokata julkaisujärjestelmässä. Työkalun kehittämisessä Kansalliskirjasto ja tutkijat toimivat yhteistyössä.

Pilottihankkeen rääätälöidyn DSpace-instanssin avulla edistetään digitoidun aineiston saavutettavuutta ja käyttöä myös ulkomailla, niin kansainvälisen tiedeyhteisön keskuudessa kuin Venäjällä asuvien suomensukuisten kielten puhujien keskuudessa. DSpace-instanssi toteutetaan myös venäjänkielisenä ja sen käyttöliittymä voidaan mahdollisesti julkaista myös muilla kielillä. Viestinnän ja tunnettavuuden edistämiseksi pilottihankkeen DSpace-instanssille luodaan räätelöity ja tunnistettava ulkoasu. Näin toteutettu julkaisujärjestelmä tukee Koneen Säätiön kieliohjelmassa esitettyjä tavoitteita avoimuuden ja vuorovaikutuksen kulttuurin edistamisestä.

Aineiston esittäminen DSpace-julkaisujärjestelmässä edistää myös aineiston löydettävyttä. Järjestelmän oma hakutoiminto tukee sekä tiedonhakua julkaisujen kuvailutiedoista että kokotekstihakua julkaisujen sisällöstä. Järjestelmään tallennetuille julkaisuille annetaan pysyvät URN-osoitteet, joiden avulla aineisto on löydettävissä, vaikka julkaisujen sijainti joskus muuttuisi. Pilottihankkeen aineisto saatetaan avoimeen käyttöön myös tulevan Kansallisen digitaalisen kirjaston (KDK) asiakasliittymän ja vastaavien kansainvälisten hakupalveluiden kautta. Aineistot ovat löydettävissä myös Googlella ja muilla yleisillä hakukoneilla.

TEKIJÄNOIKEUDET

Pilottihankkeessa on otettava huomioon sekä suomalaiset että venäläiset tekijänoikeuslainsäädännöt, koska digitointi tehdään Venäjällä ja aineisto asetetaan käyttöön Suomessa. Koska suurin osa digitoitavasta aineistosta kuuluu tekijänoikeuslainsäädännön piiriin, niin aineiston digitointia ja luovuttamista koskevia edellytyksiä on edelleen selvitettävä yhteistyössä Venäjän Kansalliskirjaston ja National Library Resourcen kanssa.

VIESTINTÄ

Pilottihankkeen viestintä kuuluu projektipäällikön työtehtäviin yhteistyössä Kansalliskirjaston viestintäyksikön kanssa. Pilottihankkeelle luodaan oma hankesivu Kansalliskirjaston verkkosivuille. Tämän sivukokonaisuuden avulla tiedotetaan hankkeen tapahtumista tarvittavilla kielillä, vähintään suomeksi, ruotsiksi, englanniksi ja venäjäksi sekä mahdollisuuksien rajoissa myös muilla kielillä. Hankkeen käytettävissä ovat myös muut Kansalliskirjaston yleiset viestintäkanavat, kuten asiakaslehdet ja sosiaalisen median viestimet, joiden avulla voidaan kohdentaa viestintää ja tiedottamista erilaisille käyttäjäryhmille. Pilottihankkeen projektipäällikkö vastaa viestinnällisistä sisällöistä ja tietojen ajantasaisuudesta. Kesän ja syksyn aikana pilottihanketta on tarkoitus esitellä ja saatetaan tietoisuuteen myös erilaisissa kotimaisissa (esim. Kirjastoverkkopäivät) ja kansainvälisissä tapahtumissa (esim. IFLA 2012) yhteistyössä niin Koneen Säätiön kuin muiden toimijoiden kanssa. Pilottihanketta esitellään myös yksityiskohtaisemmin syksyllä järjestettävässä kick off- tapahtumassa, jonka tarkoituksena on viestiä niin tutkijayhteisölle kuin suurelle yleisölle.

JATKOHANKE

Pilottihankkeen aikana (heinäkuu 2012-joulukuu 2013) kehitetään toimintatapaa ja teknistä toteutusta, joita on tarkoitus hyödyntää muiden kieliohjelmaan liittyvien aineistojen digitoinnissa jatkohankkeessa vuosina 2014–2016. Tarkoituksena on luoda myös uusia edellytyksiä kansainväliselle yhteistyölle ja laajentaa mahdollisuuksien mukaan yhteistyötä kattamaan myös muita suomalais-ugrilaisia kieliaineistoja hallinnoivia ja niitä digitoineita muistiorganisaatioita Venäjällä ja Suomessa. Yhteistyön kehittäminen pilottihankkeen aikana luo paremmat edellytykset jatkohankkeen toteuttamiselle, erityisesti digitaalisten aineistojen metadataformaattien (METS/Alto-paketit ja RusMarc-konversio) suhteen.

* * *

YHTEYSTIEDOT JA TIEDUSTELUT

Jussi-Pekka Hakkarainen
Projektipäällikkö
Tutkimuskirjasto

Kansalliskirjasto
PL 15 (Unioninkatu 36)
00014 HELSINGIN YLIOPISTO
+358 9 191 40793
+358 50 363 9223
jussi-pekka.hakkarainen@helsinki.fi

