

Kielipankin johtoryhmän kokouksen pöytäkirja 5.4.2006

Paikalla: [EeroHyvonen](#), [KimmoKoskenniemi](#), [ManneMiettinen](#), [AnnaMauranen](#), [PirkkoNuolijarvi](#), [StefanWerner](#), [AnssiYliJyra](#) (siht)

1. Kokous alkoi 12.05

2. CSC:n kuulumiset

- Lemmien käyttäjiä yli 400, suurin CSC:n käyttäjäryhmä
- Vuonna 2005 harrastettu kansainvälistä verkostoitumista osallistumalla FSMNLP2005-tapahtuman organisointiin
- CSC:ssä valmisteltu viime kuukausina KitWiki-palvelupilottia <http://forums.csc.fi/kitwiki/>
- Finnish Broadcast Corpus täydennysosa 2 toimitettu ja asennettu CSC:n koneille. Tästä on tulossa lisätietoa piahkoin kun YLE:n sopimus on ajan tasalla.

3. Keskusteltiin kieliaineistojen lisenssimalleista

- konsortiosopimukset mahdollistavat kalliimpien, mutta harvojen tahojen tarvitsemien aineistojen hankkimisen keskitetysti - LDC ja LREC aineistoja ei olla onnistuttu saamaan CSC:hen koska CSC ei ole aineistojen tarvitsija - Kenties eri laitosten muodostama konsortio voisi yhdessä CSC:n kanssa olla LDC:n ja LREC:in osapuoli. Selvitettävä - konsortiosopimuksessa tarvitsijat voivat osallistua kustannuksiin (konsortiosopimusta muistuttava tapausesimerkki: IAAS:n Swahili Language Manager eli SALAMA on nykyisin sijoitettu CSC:lle, mutta IAAS maksaa siihen kuuluvien ohjelmien lisenssit)
- ohjelmistojen ja korpusten keskitetyt lisenssit
(tapausesimerkki
CSC:n corpus.csc.fi-koneella on erilaisia kaupallisia kieliteknologiatyökaluja tutkijoiden käytössä)
- käyttöpalvelun tuottamissopimukset (esim. Oulun korpus on edelleen Oulun korpus, mutta käyttäjät pääsevät siihen CSC:n kautta. CSC on sopinut Oulun yliopiston kanssa siitä mitä ehtoja aineiston käyttäjiltä edellytetään ja kuka myöntää aineistoon käyttöoikeuden.)
- [KimmoKoskenniemi](#): uusien kieliaineistojen sopimukset laadittava Suomessa Kielipankkia silmällä pitäen
 - Julkisesti maksetut projektit julkiseksi
 - vältettävä eksklusiivisia lisenssejä kustantajien kanssa
 - vältettävä akateemiseen käyttöön rajoittuneita lisenssejä: esimerkki Norjasta, jossa tämä on opittu kantapään kautta
 - Saamen tutkijat esimerkkinä...
- uudeksi Kielipankin aineistoksi kelpuutetaan aineisto, jolle on
 - a. määritelty aineiston lähde riittävästi
 - b. jolle on määritelty kuka antaa lupia aineistoon
 - c. kohtuullinen kuvaus käytetystä annotaatiosta
- Aineiston annotaation asteelle ei aseteta vaatimuksia, vaan aineistoja on eri tyyppisiä:
 - a. puutarhoja, joissa on vain pitkälle analysoituja ilmauksia
 - b. arboretumeja, josta tiedetään suurinpiirtein mitä aineistossa on annotoitu
 - c. viidakkoja, jotka vastaavat eri tasoisia raakatekstejä
- käyttöoikeussopimukset

- Kielipankin käyttöluvahakemus (sitouttaa käyttäjän käyttämään CSC:llä olevia kieliresursseja määritellyllä tavalla sen jälkeen kun on saanut niihin käyttöoikeuden myöntäjältä luvan ja CSC:ltä käyttäjätunnuksen)
 - kurssitunnukset (yliopistojen opettajat voivat hakea kymmenen numeroitua tunnusta kurssia varten ja tämän jälkeen opettajat ovat vastuussa tunnusten käytöstä)
 - erillisaineistojen käyttöoikeudet (tutkimusryhmän edustaja tai kurssin opettaja voi hakea ryhmälleen aineistoja sopimalla CSC:n ja ohjattavien kanssa aineistojen käyttöehdoista)
- Kaikkien sopimustyyppien sopimusohjat pitäisi saada verkkoon
 - Aineistoja koskeva luokittelu
 - akateemisesti hyödynnettäviin
 - kieliteknologian kehittämisessä hyödynnettäviin

4. Keskusteltiin siitä miten Kielipankki voi parantaa Suomen ja lähialueiden kilpailukykyä ja kansainvälistä vaikuttavuutta

- alueen kannalta relevanttien kielten käännöskorpuukset
- suomenkieliset ontologiat ja muut relevantit leksikaaliset resurssit
- yhteiset kieliresurssit paremmin kieltemme kieliteknologian käyttöön
- sopimusmallien kautta käytäntöihin vaikuttaminen
- keskittämisen kautta kustannussäästöihin ja kilpailukykyyn

5. CSC:n näkökulmasta esiintulevia kehityssuuntia esitteli [AnssiYliJyra](#)

- a. kustannustehokkuutta synergialla ja keskittämisellä => CSC - the Finnish IT Center for Science
- b. johdonmukaisuutta työtä rahoittavilla konsortioilla => ei tehdä turhia hankintoja
- c. asiakkaan ja kielipankin ylläpidon yhteistyö toteuttamisessa => KitWikissä
- d. projektointi vs. riskit ja piilossa olevat sitoumukset => rönsyjen hallinta
- e. yhteisten suomen leksikaalisten resurssien merkitys korpusten rinnalla => esim. laajasti hyödynnettävät morfologisesti tai leksikaalisesti annotoidut korpuukset ja puupankit
- f. CSC mukana morfologiantutkimuksen verkostoitumisessa => FSMNLP-verkostot
- g. harvinaisten kieliaineistojen keskitetyt käyttöpalvelut => University of Helsinki Language Corpus Server voidaan periaatteessa siirtää CSC:hen
- h. kielen tutkijoiden standardin atk-ympäristön määrittely => corpus.csc.fi-UNIX-koneesta yhteinen palvelin
- i. kielipankin sopimusten tuotteistaminen => sopimusohjat saataville, tutkijoiden itseohjautuvuuden lisääminen
- j. kielitieteen palvelusta kielentutkimuksen palveluksi => esim. historialliset Melamnu-tekstit

6. Leksikaalisten ja muoto-opillisten resurssien kasvattaminen

- a. asiasanastot, ontologiat, [WordNet](#), [FrameNet](#):
 - [EeroHyvonen](#): suomen kielen ontologioista tulee julkisia (MIT-tyyppinen OS lisenssi), voidaan sijoittaa siis myös CSC:hen
- b. avoimen saatavuuden lisääminen.
 - Center for Open Source Software (COSS) ja erityisesti Elias Aarnio edistämässä open-source yhteistyötä
- c. Kielitoimiston sanakirjan sanalista (ei hakusana-artikkeleita) saatavissa Kotukselta Kielipankkiin

7. Yhteisen atk-ympäristön määrittely [KitWikin avulla](#)

- Demosimme alustavan [KitWiki](#)-järjestelmän piirteitä ja keskustelimme mm. sen käyttämisestä kielentutkijan UNIX-ympäristön rakentamisessa.
- Todettiin että yhteisten kieliteknologiamenetelmien ja UNIX-ympäristön rakentaminen on Kielipankin kehittämisessä keskeisen tärkeää.

8. Muut asiat

- Todettiin että Suomen Kielipankin johtoryhmän tehtävänä on antaa neuvoja ja ohjata CSC:n asiantuntijaa. Näin nämä kokoukset eivät tee varsinaisia päätöksiä, vaan orientoivat yhteistyötä.

9. Kokous päätettiin n. klo 14.

-- [AnssiYliJyra](#) - 26 Apr 2006

Pöytäkirjan ovat tarkistaneet: --pöytäkirjaa ei ole vielä tarkistettu--

Topic revision: r12 - 2006-09-04 - [AnssiYliJyra](#)

Copyright © 2008-2019 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.
Ideas, requests, problems regarding TWiki? [Send feedback](#)

