

Anssin viesti 1.

T. Pirkko

----- Forwarded message from Anssi Yli-Jyrä <yljyra@csc.fi> -----

Date: Mon, 28 Aug 2006 19:09:38 +0300 (EEST)

From: Anssi Yli-Jyrä <yljyra@csc.fi>

Reply-To: Anssi Yli-Jyrä <yljyra@csc.fi>

Subject: UHLCS datan siirto CSC:lle

To: ling@ling.helsinki.fi, pirkko.suihkonen@helsinki.fi

Pirkko,

En ole tietoinen viimeisimmästä kehityksestä, mutta oletan että ylkt.laitoksen linjauksena on korvata venus corpus.csc.fi:lla niin että venuksen ylläpitoa monipuolisena tutkimuspalvelimena voidaan alkaa vuoden vaihteen jälkeen poistamaan.

Olette näköjään saaneet nili.helsinki.fi levypalvelimen, joten UHLCS:n materiaali voi kai edelleen olla siellä, mutta kahdessa paikassa ei saisi olla muokkausmahdollisuutta. Muokkaus tapahtuisi parhaiten siellä missä on myös ohjelmistot eli CSC:llä. CSC:n ohjelmistoissa on suoritettava täydentelyä.

UHLCS-materiaalin siirtämiseksi CSC:lle olen varannut sinun käyttöösi hakemiston:

/mnt/corpus/uhlcs/

Ajatuksena on että tämä hakemisto olisi suoraan yleisen kt. laitoksen hallinnassa.

Olen selvittellyt tiedostojen siirron problematiikkaa:

<http://forums.csc.fi/kitwiki/pilot/view/KitWiki/Helpdesk:INeedMoreDiskSpace>

Mutta tämä ei ole nyt ehkä relevanttia. Nyt saattaa siirtämisessä olla parasta kysyä neuvoa ATK-keskuksesta tai Kimmolta. Meillä pitäisi olla nykyaikaisia nauha-asemia, mutta siirto voisi onnistua myös verkon yli. Siirtoformaatiksi ehdotan "tar cvfz" -muotoa, joka täällä puretaan --same-permissions asetuksella. Olin kuitenkin ymmärtävinäni Kimmolta että Lingin käyttäjäkuntaa ei siirretä CSC:lle ja ryhmissä pyritään reduktioon. Tar-tiedostosta ajattelin tuottaa myös tiedostolistan, jossa näkyy kaikki tiedostot alkuperäisine oikeuksineen, käyttäjineen ja ryhmineen. Siinä on historiallinen referenssi, mutta en tiedä miten tiedostojen käyttäjäryhmät nyt asetetaan.

Yhtenä myöhempänä ongelmana on että korpuksista tulee ainakin toistaiseksi kaksi versiota: venus ja corpus. Kuinka nämä versiot pidetään synkronissa? Kumpi on virallinen, jos jompaa kumpaa muutetaan? Voidaanko esim. laitoksen /ling/corpus-hakemisto jäädyttää

read-only-tilaan? Siirron jälkeen on satsattava siihen että laitoksen tutkijat siirtyvät vauhdilla CSC:n koneympäristöön niin että venuksen kuorma vähenee.

Esittäisin tavoitteeksi että siirto toteutetaan lokaluun alussa, mikäli corpus3.csc.fi on jo onnistuttu ajamaan sisään.

Haluasitko kommentoida tai kertoa missä mennään, niin tiedän onko minulla sama käsitys tai onko siellä jotain uutta.

Hyvää loppukesää!

Anssi

Viesti 2.
T. Pirkko

----- Forwarded message from Kielipankin yllapitaja <ling@csc.fi> -----
Date: Tue, 29 Aug 2006 11:36:29 +0300 (EEST)
From: Kielipankin yllapitaja <ling@csc.fi>
Reply-To: Kielipankin yllapitaja <ling@csc.fi>
Subject: Re: UHLCS datan siirto CSC:lle
To: suihkone@mappi.helsinki.fi

Pirkko,

> Tässä osoitteessa olevat ohjelmat pitäisi saada
> csc:lle:
> <http://www.ling.helsinki.fi/atk/sovellusohj/parsers/index.shtml>.

- * suomen morfologinen jäsennin: TW-FIN
- * suomen morfologinen jäsennin: MORFO
- * ruotsin morfologinen jäsennin: TW-SWE
- * saksan morfologinen jäsennin: TW-GER
- * englannin dependenssijäsennin EN-FDG
- * ruotsin dependenssijäsennin SV-FDG
- * suomen dependenssijäsennin FI-FDG
- * yksiselitteistävä TEXTMORFO-jäsennin
- * englannin morfosyntaktinen jäsennin: ENGCG
- * saksan morfosyntaktinen jäsennin: GERCG
- * suomen morfosyntaktinen jäsennin: FINCG
- * ruotsin morfosyntaktinen jäsennin: SWECG

KAIKKI NÄMÄ OVAT JO OLLEET MUUTAMIA VUOSI CSC:LLÄ (paitsi GERCG ja TW-GER)

mutta niistä on tiedotettu luvattoman huonosti.

Anssi

Viesti 3.
T. Pirkko

----- Forwarded message from ylijiyra@csc.fi -----
Date: Wed, 7 Feb 2007 23:04:59 +0200 (EET)
From: ylijiyra@csc.fi
Reply-To: ylijiyra@csc.fi
Subject: [Fwd: Re: UHLCS -asiaa]
To: ling@csc.fi, manne.miettinen@csc.fi

Eero ja Manne,

Ohjaan tämän nyt teille hoitavapaani vuoksi. Pirkko on kuulolla (cc).

Omat ajatukseni voisi tiivistää näin:

UHLCS on verrattavissa korpuksiin, joiden käyttöluvan myöntää jokin yhteyshenkilö joltakin laitokselta (tällaisia aineistojahan meillä on jo.), mutta käyttöluvan myötävään laitokseen saa yhteyden CSC:n

kautta. Laitos myös vastaa siitä että käyttöehdot on määritelty jossakin (en kannata lisätyötä, mutta asiaan pitää sopia neuvotteluteitse

laitosta ja CSC:tä miellyttävä ratkaisu, muuten se jää jalkapuoleksi) johon CSC voi viitata sopimusteknisesti välttämättömällä tavalla niissä asiapapereissa joita käyttäjät allekirjoittavat.

(Käyttöoikeussopimuksessa voidaan velvoittaa käyttäjä ottamaan selvää

kukin korpuksen käyttöehdoista mistä ne sitten löytyvätkin, ja samalla tulee kertoa mistä niitä voi etsiä.)

Lopputuloksena tästä on että CSC:n käyttöluvalomake riittää.

Ajatuksena tosiaan on, että aineistot ovat edelleen suoraan kielitieteen laitoksen hallinnassa, mutta Unix-ympäristö, Unix-tunnukset ja

Unix-ryhmät

sekä keskitetty käyttöluvalomake ovat CSC:n alihankintaa, eli niinkuin Manne sitä kutsuu "CSC:n tuottama aineistojen käyttöpalvelu".

CSC:n kannalta ei

ole oleellista eroa onko UHLCS:n aineistoja varten useita eri UNIX-ryhmiä vai vain yksi ryhmä, kun taas UHLCS:n kannalta ryhmiä on nimenomaan ollut melkoisesti. Tärkeää on vain että käyttöluvalomakkeessa

yksi rastittava kohta vastaa yhtä UNIX-ryhmää. Prototyypillisesti kullekin ryhmälle ja käyttöluomalomakkeen kohdalle on myös tasan yhden käyttöehdot ja näitä vastaavat aineistot jotka siis saa rastitsemalla käyttöönsä. Koska maailma on epätäydellinen, poikkeamia voi olla.

Vedän takaisin jotkut ideani siitä että Kielipankin koneella olisi "virtuaalisia alikorpussereit". Riittää kun käyttöpalvelusopimuksessa todetaan että Kielitieteen laitoksen korpusserveri on tätä nykyisin CSC:n laskentaympäristöstä (joka voi käsittää useita koneita tai jopa virtuaalisia Grid-laskentaympäristöjä) UNIX-ryhmillä rajattu looginen osuus, jonka CSC yleisen kielitieteen laitokselle tuottaa. Tästä määrittelystä on hyötyä sellaisten korpusten kohdalla joissa laitoksen lisenssi on hankittu yhtä konetta (esim. edesmenneitä Waltaria tai Polvaa) silmällä pitäen. Juristien voi sitten vain olettaa ymmärtävän että jotkut asiat maailmassa kehittyvät teknisesti erinäköiseksi, mutta oleellinen ratkaisujen mieli ei. Kun vanha tekninen todellisuus katoaa alta, on tilalle rakennettava sanakäänteitä joilla viitataan nykyiseen toteutustekniikkaan (jossa rauta on keskitetympi).

Manne ja Eero Vitie jatkavat asian käsittelyä, mutta nämä ovat minun ajatuksiani tällä erää, kun nimeni kerran mainittiin.

Minun oli pakko kertoa näkemykseni ettei tässä synny mitään sekaannuksia kommunikointikatkon vuoksi.

terveisin
Anssi

----- Original Message -----

Subject: Re: UHLCS -asiaa
From: suihkone@mappi.helsinki.fi
Date: Wed, February 7, 2007 18:52
To: "Hanna Westerlund" <hmwester@cc.helsinki.fi>
Cc: ylijyra@csc.fi
pirkko.suikonen@helsinki.fi

Hei,

Kiitos viestistä. Hyvä, että viet asiaa eteenpäin. Olen ymmärtänyt, että Anssi tulee hoitamaan myös muiden aineistojen käyttöluvat. Luvan saamiseksi tarvitaan suositus, kuten on ennenkin tarvittu. UHLCS:n sivuilla on tietoa henkilöistä. Kyllä näissä minunkin mielestäni tarvitaan tieto siitä, mihin lupaa

ja sen avulla mahdollisuutta päästä käsiksi aineistoihin käytetään. Käyn UHLCS:n sivuilla mainitut nimet läpi kevään aikana ja päivitän mahdolliset muutokset. Jos siis Anssi hoitaa kaikki käyttöluvut, silloin ei käsittäkseni tarvita mitään meidän käyttöluomalakkeita, paitsi siis täydentämään CSC:n lomaketta. Voisiko CSC:n lomakkeeseen liittää tämän puuttuvan tiedon?

Tiedosto

-rw-rw-r-- 1 suihkone ykiel 1192 Aug 27 21:07 csc-info.html on avoin webisti-ryhmälle. Voit linkittää lisätiedot siihen. Minä linkitän sitten kirjoittamasi tiedoston myös muihin tarpeellisiin "laatikoihin" UHLCS:n kotisivulla.

Tuo

"> The Computer Application Form and a letter of motivation": riittääkö ehkä: "recommendation"?

Lähetän tämän viestin myös Anssille, niin että hänkin tietää missä mennään.

Anssi: katsotko sitä UHLCS:n sivulla olevaa käyttöluulahakemusta. Siinä olevat tiedot käyvät monelta osin yksin sen lomakkeen kanssa, joka on allekirjoitettu, kun korpukset on otettu korpusten omistajilta.

Parhain terveisin,

Pirkko

PS: Olen samaa mieltä, että CSC:n käyttöluupien löytämiseksi tarvitaan kehittyntä jäljitustekniikkaa.

Quoting Hanna Westerlund <hmwester@cc.helsinki.fi>:

>

> Hei vaan Pirkko,

>

> Laitokselle saapuu vielä sellaisia UHLCS-hakemuksia, joissa on haettu

> CSC:llä olevien aineistojen käyttöluupaa. Jos kotisivulla olevaa tekstiä

> muutettaisiin, jäisivät nämä turhat hakemukset ehkä pois.

>

> Olen vähän pohtinut, millä tavalla tiedon laittaisi UHLCS-sivulle.

>

> <http://www.ling.helsinki.fi/uhlcs/agreements/computer-account-application-form.html>

> Olisiko alla oleva sinusta mahdollinen luonnos tekstiksi?

>

> Tavoitteena on ohjata kaikki CSC:n aineistoja tarvitsevat hakemaan lupaa

>

- > suoraan sieltä. Onneksi kaikki CSC:n sivut löytyvät nykyisin myös
- > englanniksi - tiedon löytäminen tosin vaatii aika lailla työtä ja
- > laitoinkin siksi linkit mukaan tekstiin. Koska laitoksen palvelimet
- > vaativat nykyisin cc-luvan, muutin myös ko. lupalomakkeen tiedot tähän.
- > cc-lupahakemuksessa ei kysytä, mihin tarkoitukseen lupaa haetaan, joten
- > jonkinlainen selitys täytyy varmaankin pyytää erikseen.
- >
- > Terveisiä
- > Hannalta
- >
- >
- > ****
- >
- > Computer account applications
- >
- > Part of the UHLCS data has been moved to the Finnish Language Bank hosted
- >
- > by CSC Finnish IT centre for science in the beginning of 2007. This data
- >
- > includes corpora on Finnish, Swedish, English, German and Swahili. For
- > more detailed information for corpora and, see:
- >
- > http://www.csc.fi/english/research/software/index_html#field7
- >
- > Computer Accounts for these resources are applied from CSC. More
- > information is available at:
- >
- > http://www.csc.fi/english/customers/new_customer
- >
- > The Finnish Language Bank Application Form for Universities:
- >
- > <http://www.csc.fi/english/customers/university/useraccounts/languagebank>
- >
- >
- > Other linguistic resources than the above mentioned are currently hosted
- >
- > by the University of Helsinki. The use of the UHLCS is restricted to
- > concern (JATKUUN KUTEN NYK. SIVULLA)
- >
- > Computer Application Form (PDF)
- > <http://www.helsinki.fi/atk/luvat/applform.pdf>
- >
- > (SUOMENKIELISTÄ cc-LOMAKETTA EI EHKÄ TARVITA?)
- >
- > The Computer Application Form and a letter of motivation for use is asked
- >

> to be sent to the following address
>
> (OSOITE OK)
>
> ****
>
> --
> Hanna Westerlund
> Koordinaattori
> Kieliteknologian opetuksen verkosto (KIT)
> Yleisen kielitieteen laitos
> PL 9 (Siltavuorenpenger 20 A, 3. krs huone 314)
> 00014 HELSINGIN YLIOPISTO
> puhelin: (09) 191 29326, faksi: (09) 191 29307
> sähköposti: hanna.westerlund@helsinki.fi
>

Viesti 4.

T. Pirkko

----- Forwarded message from Kielipankin yllapitaja <ling@csc.fi> -----

Date: Tue, 20 Feb 2007 09:44:38 +0200 (EET)

From: Kielipankin yllapitaja <ling@csc.fi>

Reply-To: Kielipankin yllapitaja <ling@csc.fi>

Subject: Re: UHLCS -asiaa

To: suihkone@mappi.helsinki.fi

On Thu, 8 Feb 2007 suihkone@mappi.helsinki.fi wrote:

> Hei,
>
> Kiitos viestistä.
> UHLCS:n korpuksen käyttöehdot on määritelty lupalomakkeessa
>
> (<http://www.ling.helsinki.fi/uhlcs/agreements/computer-account-application-form.html>).
> Niiltä osin lupaehdot säilyvät varmasti ennallaan.
> Lisäksi kunkin korpuksen kohdalla on mainittu yhteyshenkilö,
> jolta voidaan käyttöoikeus tarvittaessa varmistaa.
>
> Parhain terveisin,
> Pirkko
>

Hei! Vahvistan vielä Anssin näkemyksen siitä, että aineistot voidaan sijoittaa CSC:lle.

> Quoting ylijyra@csc.fi:
>
>> Eero ja Manne,
>>
>> Ohjaan tämän nyt teille hoitavapaani vuoksi. Pirkko on kuulolla (cc).
>>
>> Omat ajatukseni voisi tiivistää näin:
>>
>> UHLCS on verrattavissa korpuksiin, joiden käyttöluvan myöntää
>> jokin yhteyshenkilö joltakin laitokselta (tällaisia aineistojahan
>> meillä on jo.), mutta käyttöluvan myötävään laitokseen saa >>
>> yhteyden CSC:n kautta. Laitos myös vastaa siitä että käyttöehdot
>> on määritelty jossakin (en kannata lisätyötä, mutta asiaan pitää
>> sopia neuvotteluteitse laitosta ja CSC:tä miellyttävä ratkaisu,
>> muuten se jää jalkapuoleksi) johon CSC voi viitata
>> sopimusteknisesti välttämättömällä tavalla niissä asiapapereissa
>> joita käyttäjät allekirjoittavat. (Käyttöoikeussopimuksessa
>> voidaan velvoittaa käyttäjä ottamaan selvää kukin korpuksen
>> käyttöehdoista mistä ne sitten löytyvätkin, ja samalla tulee
>> kertoa mistä niitä voi etsiä.)

CSC voi hoitaa UHLCS:n luvat ottamalla vastaan hakemukset ja lähettämällä tiedustelun oikeuksien myöntämisestä aineistojen vastuuhenkilöille. Lupaa ei myönnetä ilman vastuuhenkilön puolta.

Aineistojen käyttöehdot voidaan sisällyttää hakemuslomakkeeseen ongelmitta. Myös vaaditut henkilö- ja yhteystiedot ovat mielestäni melko lailla samat CSC:n ja Yleisen kielitieteen laitoksen atk-lupahakemuslomakkeissa. CSC:n lomakkeessa vaaditaan seuraavat tiedot:

- etu- ja sukunimi,
- virka tai toimi,
- organisaatio,
- postiosoite, -numero ja -toimipaikka,
- puhelinnumero,
- sähköpostiosoite,
- tieteellinen ohjaaja ja hänen yhteystietonsa (opiskelijoilta),
- käyttötarkoituksen laatu ja
- käyttötarkoitus.

Tieteellinen ohjaajahan vastaa käytännössä suosittelijaa, ja tämä kenttä voidaan vaatia opiskelijoiden lisäksi myös UHLCS:n aineistoihin osoitetuissa hakemuksissa.

- >> Lopputuloksena tästä on että CSC:n käyttöluomalomake riittää.
- >> Ajatuksena tosiaan on, että aineistot ovat edelleen suoraan
- >> kielitieteen laitoksen hallinnassa, mutta Unix-ympäristö,
- >> Unix-tunnukset ja Unix-ryhmät sekä keskitetty käyttöluomalomake
- >> ovat CSC:n alihankintaa, eli niinkuin Manne sitä kutsuu "CSC:n
- >> tuottama aineistojen käyttöpalvelu".

Käytännössä siis CSC ylläpitää laiteympäristöä ja käyttäjähallintoa, mutta Yleisen kielitieteen laitos (tai muu aineiston koostaja) voi vapaasti määrätä aineistoista ja niihin myönnettyistä oikeuksista. Yhteyshenkilöinä CSC:n puolelta toimivat Kielipankin ylläpito (ling@csc.fi) ja Käyttäjähallinto (usermgr@csc.fi).

Eero Vitie

Viesti 5.

T. Pirkko

----- Forwarded message from suihkone@mappi.helsinki.fi -----

Date: Tue, 20 Feb 2007 10:08:00 +0200

From: suihkone@mappi.helsinki.fi

Reply-To: suihkone@mappi.helsinki.fi

Subject: Re: UHLCS -asiaa

To: Kielipankin ylläpitaja <ling@csc.fi>

Hei,

Kiitos viestistä.

Kaikki tuntuu olevan kunnossa.

Päivitan yhteyshenkilöitä koskevat tiedot ja tarkastan vielä kerran aineistojen esittelyjä koskevat tiedot.

Toivon, että voin tehdä sen viimeistään huhtikuussa. Sopiiko?

UHLCS:ssa on joitakin aineistoja, jotka eivät ole julkisesti käytössä, koska niistä puuttuu aineistojen omistajan allekirjoittama lupa. Aineistot ovat siis olleet jonkun tutkijan yksityiskäytössä.

Niistäkin pitää tehdä luettelo, ja niiden käytöstä on sovittava ko. tutkijan kanssa.

Otan yhteyttä, jos asiaan liittyy erityisjärjestelyjä.

Parhain terveisin,

Pirkko

Quoting Kielipankin yllapitaja <ling@csc.fi>:

> On Thu, 8 Feb 2007 suihkone@mappi.helsinki.fi wrote:

>

>> Hei,

>>

>> Kiitos viestistä.

>> UHLCS:n korpusten käyttöehdot on määritelty lupalomakkeessa

>>

>

(<http://www.ling.helsinki.fi/uhlcs/agreements/computer-account-application-form.html>).

>> Niiltä osin lupaehdot säilyvät varmasti ennallaan.

>> Lisäksi kunkin korpuksen kohdalla on mainittu yhteyshenkilö,

>> jolta voidaan käyttöoikeus tarvittaessa varmistaa.

>>

>> Parhain terveisin,

>> Pirkko

>>

>

> Hei! Vahvistan vielä Anssin näkemyksen siitä, että aineistot voidaan

> sijoittaa CSC:lle.

>

>> Quoting ylijyra@csc.fi:

>>

>>> Eero ja Manne,

>>>

>>> Ohjaan tämän nyt teille hoitavapaani vuoksi. Pirkko on kuulolla

> (cc).

>>>

>>> Omat ajatukseni voisi tiivistää näin:

>>>

>>> UHLCS on verrattavissa korpuksiin, joiden käyttöluvan myöntää

>>> jokin yhteyshenkilö joltakin laitokselta (tällaisia aineistojahan

>>> meillä on jo.), mutta käyttöluvan myöntävään laitokseen saa >>

>>> yhteyden CSC:n kautta. Laitos myös vastaa siitä että käyttöehdot

>>> on määritelty jossakin (en kannata lisätyötä, mutta asiaan pitää

>>> sopia neuvotteluteitse laitosta ja CSC:tä miellyttävä ratkaisu,

>>> muuten se jää jalkapuoleksi) johon CSC voi viitata

>>> sopimusteknisesti välttämättömällä tavalla niissä asiapapereissa

>>> joita käyttäjät allekirjoittavat. (Käyttöoikeussopimuksessa

>>> voidaan velvoittaa käyttäjä ottamaan selvää kukin korpuksen

>>> käyttöehdoista mistä ne sitten löytyvätkin, ja samalla tulee

>>> kertoa mistä niitä voi etsiä.)

>

>

> CSC voi hoitaa UHLCS:n luvat ottamalla vastaan hakemukset ja

- > lähettämällä tiedustelun oikeuksien myöntämisestä aineistojen
- > vastuuhenkilöille. Lupaa ei myönnetä ilman vastuuhenkilön puolta.
- >
- > Aineistojen käyttöehdot voidaan sisällyttää hakemuslomakkeeseen
- > ongelmitta. Myös vaaditut henkilö- ja yhteystiedot ovat mielestäni
- > melko lailla samat CSC:n ja Yleisen kielitieteen laitoksen
- > atk-lupahakemuslomakkeissa. CSC:n lomakkeessa vaaditaan seuraavat
- > tiedot:
- >
- > - etu- ja sukunimi,
- > - virka tai toimi,
- > - organisaatio,
- > - postiosoite, -numero ja -toimipaikka,
- > - puhelinnumero,
- > - sähköpostiosoite,
- > - tieteellinen ohjaaja ja hänen yhteystietonsa (opiskelijoilta),
- > - käyttötarkoituksen laatu ja
- > - käyttötarkoitus.
- >
- > Tieteellinen ohjaajahan vastaa käytännössä suosittelijaa, ja tämä
- > kenttä voidaan vaatia opiskelijoiden lisäksi myös UHLCS:n aineistoihin
- > osoitetuissa hakemuksissa.
- >
- >>> Lopputuloksena tästä on että CSC:n käyttölupalomake riittää.
- >>> Ajatuksena tosiaan on, että aineistot ovat edelleen suoraan
- >>> kielitieteen laitoksen hallinnassa, mutta Unix-ympäristö,
- >>> Unix-tunnukset ja Unix-ryhmät sekä keskitetty käyttölupalomake
- >>> ovat CSC:n alihankintaa, eli niinkuin Manne sitä kutsuu "CSC:n
- >>> tuottama aineistojen käyttöpalvelu".
- >
- > Käytännössä siis CSC ylläpitää laiteympäristöä ja käyttäjähallintoa,
- > mutta Yleisen kielitieteen laitos (tai muu aineiston koostaja) voi
- > vapaasti määrätä aineistoista ja niihin myönnettyistä oikeuksista.
- > Yhteyshenkilöinä CSC:n puolelta toimivat Kielipankin ylläpito
- > (ling@csc.fi) ja Käyttäjähallinto (usermgr@csc.fi).
- >
- > Eero Vitie