

Tuotokset: aineistokorpukset ja kirjoitukset

HPD-webinaari 19.1.2021

Mona Lehtinen & Juho Inkinen

Korpuukset

- KK toimitti CSC:lle aineistoa tutkimusta varten oman rajoitetun aineiston korpuksen
 - Vapaakappaleita, kirjojen esittelytekstejä
 - Tekijänoikeussyistä aineistot käsitelty niin, ettei Kansalliskirjaston ulkopuolelle siirry teosta (sekoittaminen)
- Lisäksi julkinen [annif-corpora](#) GitHubissa
 - Finna-aineisto
 - JYX-gradut
 - Kysy kirjastonhoitajalta

Korpukset, Finna

- Algoritmien kouluttamiseen tarkoitettu kolmikielinen (fi, sv, en) aineisto
- Muodostettu Finna.fi-hakupalvelun viitetietueista, joilla on Yleisestä suomalaisesta ontologiasta (YSO) poimitut asiasanat.

Korpukset, JYX

- Kokotekstiaineisto, joka on muodostettu Jyväskylän yliopiston julkaisuarkistosta poimituista suomen-, ruotsin- ja englanninkielisistä opinnäytteistä (gradut ja väitöskirjat).
- Aineistoa voidaan käyttää sekä algoritmien kouluttamiseen että arviointiin.

Korpukset, Kysy kirjastonhoitajalta

- Kokotekstiaineisto, joka on muodostettu Kysy kirjastonhoitajalta -palvelun suomenkielisistä kysymys-vastauspareista. [Julkaistu 4/2018](#).
- Yhteensä 3150 dokumenttia, aineistoa voidaan käyttää sekä algoritmien kouluttamiseen että arviointiin.

Kirjoitukset

1. Blogikirjoitus ([suomeksi](#) ja [englanniksi](#))
2. Sisällönkuvailun automatisoinnin haasteita ja ratkaisuja kulttuuriperintöorganisaatiossa
 - Whitepaper, tekninen kuvaus automaattisen kuvailun palvelusta
3. Automaattisen kuvailun palvelun integroiminen Kansalliskirjaston järjestelmäkokonaisuuteen - tietovirrat ja prosessit
 - Proof of Concept, käsittelee automaattisen kuvailun käyttöönottoa Kansalliskirjastossa

Kirjoitukset

- Kummatkin raportit (2 ja 3, whitepaper ja PoC) julkaistaan Kansalliskirjaston [Raportteja ja selvityksiä](#)-verkkójulkaisusarjassa



www.kansalliskirjasto.fi

Kiitos!

