

Ítarupplýsingar frá Almannaróm um afhendingu málgagna.

Ég ætla setja hér inn smá Q&A kafla sem ætti vonandi að snerta á sem flestum tæknilegum spurningum:

Hvað verður um heimildirnar sem safnast í þessu verkefni?

o Öll þau textagögn sem safnast verða afhent samstarfsaðila okkar, Stofnun Árna Magnússonar í íslenskum fræðum, sem undirbýr, hreinsar, framkvæmir málfræðilega greiningu, brýtur efnið niður og staðlar á rafrænu formi.

o Á endanum verður svo til málheild fyrir átakið sem gefin er út undir opnum leyfum sem kallast CC by 4.0 (Creative Commons Attribution 4.0 International). Það þýðir að öllum sé frjálst að nota málheildina.

Verða viðkvæmar upplýsingar sjálfkrafa hreinsaðar úr skjölum?

o Öll gögn sem afhent eru í verkefnið verða hreinsuð að því leiti að fjarlægð eru HTML-merki, vefslóðir, kóða og aðra hluti sem falla ekki undir náttúrulegan íslenskan texta.

*o **Ekki** er átt við efni textagagna og þau eru ekki hreinsuð af því sem talist gæti viðkvæmar upplýsingar. Því er sérstaklega mikilvægt að sendandi staðfesti að skjöl/textagögn sem látin eru af hendi í átakið innihaldi ekki viðkvæmar upplýsingar af neinum toga.*

Hverjir munu hafa aðgang að gagnahirslum sem heimildir verða geymdar í?

o Aðeins starfsmenn Stofnunar Árna Magnússonar í íslenskum fræðum munu hafa tæknilegan aðgang að þeim gögnum sem afhent verða í verkefnið og aðeins í þeim tilgangi að vinna þau eins og er lýst hér að framan. Tilgangurinn með gagnavinnslnunni er að bæta orðaforða og málnotkun íslenskra fyrirtækja í okkar stóra orðasafn og auka þannig gæði þess og fjölbreytileika.

Verður hægt að endurgera og leita í þeim skjölum sem ég gef af hendi?

o Nei, þegar textagögn eru unnin fyrir málheildir eru þau bútuð niður svo aðeins takmarkað samhengi er hægt að sækja út frá hverju leitarskilyrði fyrir sig.

Get ég óskað eftir því að veittum heimildum sé eytt úr málheildinni eftir á?

o Risamálheild er uppfærð árlega og ef óskað er eftir því er hægt að fjarlægja gögn úr málheildum á nýjum uppfærslum en þau verða þó áfram í eldri útgáfum. Ekki er þó hægt að afturkalla öll gögn úr núverandi eða eldri útgáfum á málheildum.

Við munum svo opna gátt sem hægt verður að nota til að gefa heimildir í átakið á mánudaginn, endilega hafðu samband ef það vakna aðrar spurningar

Mbk.

Valgeir Haukdal

Tæknilegur Verkefnastjóri

