

Aukið við safn sendibréfa frá f.hl. nítjándu aldar

07. júlí 2015

Heimir van der Feest Viðarsson

Í sendibréfasafninu (<http://brefasafn.arnastofnun.is>) er eins og stendur býsna lítið af bréfum frá f.hl. 19. aldar, sbr. eftirfarandi yfirlit sem er gróft áætlað (ath. að XML-hylkin eru talin með, svo tölurnar eru ofáætlaðar):

Tímabil	Orðafjöldi
1800-1810	0
1810-1820	3.710
1820-1830	19.501
1830-1840	16.833
1840-1850	30.323
1850-1860	127.584
1860-1870	197.691
1870-1880	177.486
1880-1890	146.291
1890-1900(+)	132.102

Til þess að fá jafnari dreifingu yfir öldina („stoppa í götin“) verður texti eftirtalinna útgáfna gerður aðgengilegur rafrænt:

1. *Biskupinn í Görðum*, bls. 9-326. Finnur Sigmundsson (útg.). Rvk., 1959.
2. *Geir biskup góði í vinarbréfum 1790-1823*, bls. 11-212. Finnur Sigmundsson (útg.). Rvk., 1966.
3. *Hafnarstúdentar skrifa heim*, bls. 9-109. Finnur Sigmundsson (útg.). Rvk., 1963.
4. *Húsfreyjan á Bessastöðum*, bls. 1-241. Bréf Ingibjargar Jónsdóttur til bróður síns, Gríms amtmanns. Finnur Sigmundsson (útg.). Rvk. 1946.
5. *Konur skrifa bréf*, bls. 9-148. Sendibréf 1797-1907. Finnur Sigmundsson (útg.). Rvk., 1961.
6. *Sendibréf frá íslenskum konum 1784-1900*, bls 1-99. Finnur Sigmundsson (útg.). Rvk. 1952.
7. *Bjarni Thoranensen: Bréf*, I.-II. bindi. Jón Helgason (útg.), Hið íslenska fræðafélag, Rvk. 1986. [Bréfum á dönsku sleppt.]

Þessi viðbót er um 400.000 orð.

Aðeins voru skönnuð bréf sem lenda innan tímabilsins ca 1800-1850. Ljósmyndirnar voru síðan OCR-ljóslesnar með Tesseract-OCR.¹ Útkoman er yfirleitt mjög góð en þó svolítið misjöfn eftir útgáfum. Verst kom út elsta útgáfa útg. Finns, *Sendibréf frá íslenskum konum 1784-1900*. Ætlunin er að textanum verði skipt upp í bréf skv. nafni bréfitara, ártali og dagsetningu eins og bréfunum í bréfasafninu en þetta hefur einvörðungu verið gert fyrir tímabilið 1800-1825 eins og er.

1 Tesseract-OCR. 2014. Útgáfa 3.03. Google. <<https://code.google.com/p/tesseract-ocr/>>.

Athugið að í textunum leynast villur. Æskilegt er því að textinn verði einnig leiðréttur með leiðréttingartóli á borð við Skramba. Þeir textar eftir útg. Finns sem hefur verið skipt upp í bréf (sbr. hér á undan) hafa verið leiðréttir vélrænt að nokkru leyti – þ.e. algengustu villur lagaðar. Minna hefur verið átt við bréf Bjarna Thorarensens. Athugið enn fremur að síðastnefnda útgáfan er stafrétt, ólíkt öllum hinum.