

IceAlign

Search Tool for the Icelandic Parallel Corpus

Egill Anton Hlöðversson
University of Iceland
Language resources for software development and research

29. Mars 2019

Code





https://github.com/egillanton/icealign

Dataset

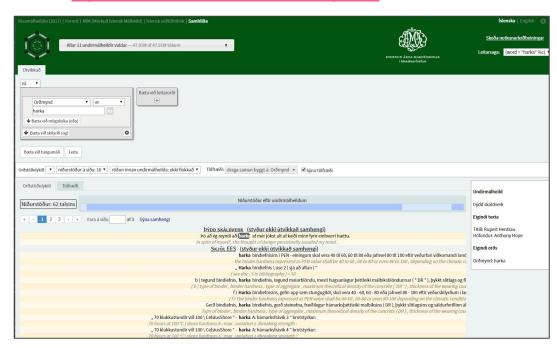
- Icelandic Parallel Corpus (IPC) :
 - http://malfong.is/?pg=samhlida
- 11 Corpora
- Icelandic English paragraphs paired together.
- .tmx files
- 3.561.789 line pairs
- Largest Corpus: EES Documents, 1.701.172
- Smallest Corpus: Statistics Iceland (from website), 2.288

Dataset

Name	Date modified	Туре	Size
baekur.tmx	2/26/2019 5:31 PM	TMX File	4,726 KB
biblian.tmx	2/26/2019 5:31 PM	TMX File	15,540 KB
ees.tmx	2/26/2019 5:32 PM	TMX File	596,666 KB
ema.tmx	2/26/2019 5:32 PM	TMX File	124,162 KB
eso.tmx	2/26/2019 5:32 PM	TMX File	5,351 KB
fornritin.tmx	2/26/2019 5:32 PM	TMX File	5,993 KB
hagstofan.tmx	2/26/2019 5:32 PM	TMX File	1,147 KB
kde4.tmx	2/26/2019 5:32 PM	TMX File	8,869 KB
opensubtitles.tmx	2/26/2019 5:32 PM	TMX File	300,458 KB
readme.txt	2/26/2019 5:46 PM	TXT File	1 KB
tatoeba.tmx	2/26/2019 5:32 PM	TMX File	1,506 KB
ubuntu.tmx	2/26/2019 5:32 PM	TMX File	1,667 KB

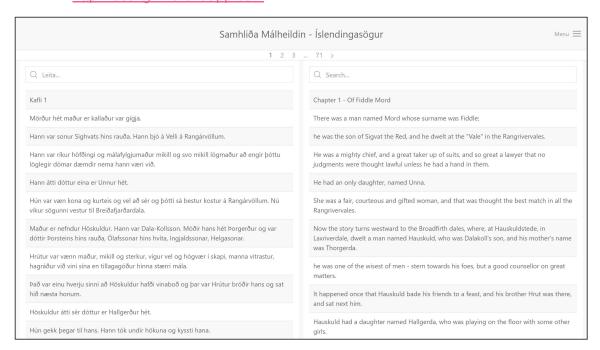
Korp

https://málheildir.árnastofnun.is/?mode=parallel



IceAlign

http://icealign.herokuapp.com



Thank You!

