

CS4642 - Data Mining & Information Retrieval

Parliament Members Search Engine

170326U

Data set

Data for the search engine was about the members of the parliament in Sri Lanka. It was scrapped from [ශ්‍රී ලංකා පාර්ලිමේන්තුව - මන්ත්‍රීවරුන්ගේ නාමාවලිය](#) website. Beautifulsoup, which was a python module used for web scraping. There were 8 fields scraped. Below is an example of the data scraped from the above website.

Name	අරවින්ද්‍ය කුමාර්
Date of Birth	1954-11-17
Marital status	විවාහක
Religion	හින්දු
Party	සමගි ජන බලවේගය
District	බදුල්ල
Telephone	0777727472
Email	aravindh_k@parliament.lk

Indexing

Elastic search was used for implementing the search engine. It was hosted in the elastic cloud. Kibana is used for the front end of the application. For indexing standard tokenizer which was provided from the elastic search was used.

Supported Queries

ආගම : "බෞද්ධ"

කොළඹ

පක්ෂය : "ජාතික ජන බලවේගය" or පක්ෂය : "ශ්‍රී ලංකා නිදහස් පක්ෂය"

දිස්ත්‍රික්කය : "ගම්පහ" and පක්ෂය : "ශ්‍රී ලංකා පොදුජන පෙරමුණ"

Git Repository

The codes were uploaded to the below GitHub repository with the proper readme.

<https://github.com/Hiruna-Kumara/Parliament-Members-Search-2021>