**BIRDS**

* Made a list of uniquely named birds by merging different datasets collected kaggle and GitHub.
* Listed out some scrapable websites which contains major attributes like ebird, eol, iucn etc….
* Collected data from different Sites such as [Ebird.org](https://ebird.org/home) (5526 birds), [Eol.org](https://eol.org/) (6571 birds) , Wikidata [[Wikipedia.org](https://www.wikipedia.org/)] (11060 birds) and [dibird.com](https://dibird.com/) (10772 birds).
* Scraped attributes like breeding region, Old Latin Name etc of birds using Tools BeautifulSoup and Selenium from the website [Dibird](https://dibird.com/). ([Code](https://github.com/Mahanth-Maha/indicwiki_birds/blob/main/curDir/scrape_data/Dibird_Scrape_Optim.py))
* Scraped different attributes of birds using Tools BeautifulSoup from the website [ebird](https://ebird.org/home). ([Code](https://github.com/Mahanth-Maha/indicwiki_birds/blob/main/curDir/scrape_data/Ebird_Scraper_Optim.py))
* Scraped different attributes like locomotion, Habitat, auditory system etc of birds using Tools BeautifulSoup from the website [eol](https://eol.org/). ([Code](https://github.com/Mahanth-Maha/indicwiki_birds/blob/main/curDir/scrape_data/Eol_Scraper_Optim.py))
* Scraped different attributes of birds using Tools Selenium from the website [iucn](https://www.iucn.org/). ([Code](https://github.com/Mahanth-Maha/indicwiki_birds/blob/main/curDir/scrape_data/Iucn_Scraper_Optim.py))
* Scraped taxonomy of birds and images of the birds using Tools like Selenium and BeautifulSoup. ([Code](https://github.com/Mahanth-Maha/indicwiki_birds/blob/main/curDir/scrape_data/scrape_wikidata.ipynb))
* Merged all datasets and performed preprocessing and data cleaning.
* The merged data resulted a ***12641*** unique birds and dropped all attributes which are less than *25%* .
* Merged data consists of 25+ attributes which have 50% and more data.
* Created a Taxonomy for the birds from the data collected : [Taxonomy](https://docs.google.com/document/d/1KSZGcSCEoy0C-6YFPwkQq51Gvm3QG5dY/edit?usp=sharing&ouid=117054022486472891083&rtpof=true&sd=true)
* Sweetviz report html : [Report](https://drive.google.com/file/d/1AgE3-x_BhWYijd1_udZsMBX8DnuUDpoK/view?usp=sharing)
* Translated all the columns of the dataset into telugu. (except the descriptions from eol and ebird columns) [Datase](https://docs.google.com/spreadsheets/d/1fBewq5cluIa7SxF_H_wOWGLRhS4qswmH/edit?usp=sharing&ouid=112798495723518296608&rtpof=true&sd=true)t using Anuvaad and GoogleTranslators.
* Worked on a few translation modules to translate the text in the description and proceeded with deep translate.
* After data cleaning, translated all the descriptions using deep translate.
* Made a [sweetwiz report](https://drive.google.com/file/d/1v9a7m6jZfIIJksEkmgrHhBELN5WFHyAN/view?usp=sharing) of the [new dataset](https://docs.google.com/spreadsheets/d/12AxHl5GfiWJQxHGGAucbj3wwV2drLEuM/edit?usp=sharing&ouid=112798495723518296608&rtpof=true&sd=true) and analysed the data.
* Worked on jinja templates and completed the code for introduction part and some other subtopics including randomisation of sentences.
* Started rectification of errors & corrections manually using a chrome extensions to compare the translation/transliteration of a word in different translators. [Extension](https://chrome.google.com/webstore/detail/translation-comparison/kicpmhgmcajloefloefojbfdmenhmhjf)
* Written around 2 articles on different birds, to get an idea on how an article actually looks. [Siao pitta](https://tewiki.iiit.ac.in/index.php?title=%E0%B0%B8%E0%B0%BF%E0%B0%AF%E0%B0%BE%E0%B0%B5%E0%B1%8D_%E0%B0%AA%E0%B0%BF%E0%B0%9F%E0%B1%8D%E0%B0%9F),[mascarene parrot](https://tewiki.iiit.ac.in/index.php?title=%E0%B0%AE%E0%B0%B8%E0%B1%8D%E0%B0%95%E0%B0%B0%E0%B1%87%E0%B0%A8%E0%B1%8D_%E0%B0%9A%E0%B0%BF%E0%B0%B2%E0%B1%81%E0%B0%95).
* Final Dataset: [Dataset](https://docs.google.com/spreadsheets/d/1mExPkrOJtYCT4Cx-I8WsxIRGncrRP4Ua/edit?usp=sharing&ouid=117054022486472891083&rtpof=true&sd=true)
* Written Jinja Template code for the articles including randomisation of sentences .

Jinja template Code: [Code](https://github.com/Mahanth-Maha/indicwiki_birds/blob/main/curDir/Template/birds.j2)

* Rendered 3-4 articles to test the jinja template code ( [అబాట్ బూబీ](https://tewiki.iiit.ac.in/index.php?title=%E0%B0%85%E0%B0%AC%E0%B0%BE%E0%B0%9F%E0%B1%8D_%E0%B0%AC%E0%B1%82%E0%B0%AC%E0%B1%80) , [అబెరాంట్ బుష్-వర్బ్లర్](https://tewiki.iiit.ac.in/index.php?title=%E0%B0%85%E0%B0%AC%E0%B1%86%E0%B0%B0%E0%B0%BE%E0%B0%82%E0%B0%9F%E0%B1%8D_%E0%B0%AC%E0%B1%81%E0%B0%B7%E0%B1%8D-%E0%B0%B5%E0%B0%B0%E0%B1%8D%E0%B0%AC%E0%B1%8D%E0%B0%B2%E0%B0%B0%E0%B1%8D) , [ఆఫ్రికన్ పారడైజ్-ఫ్లైక్యాచర్](https://tewiki.iiit.ac.in/index.php?title=%E0%B0%86%E0%B0%AB%E0%B1%8D%E0%B0%B0%E0%B0%BF%E0%B0%95%E0%B0%A8%E0%B1%8D_%E0%B0%AA%E0%B0%BE%E0%B0%B0%E0%B0%A1%E0%B1%88%E0%B0%9C%E0%B1%8D-%E0%B0%AB%E0%B1%8D%E0%B0%B2%E0%B1%88%E0%B0%95%E0%B1%8D%E0%B0%AF%E0%B0%BE%E0%B0%9A%E0%B0%B0%E0%B1%8D) *)*
* Rectified the errors in article generated and corrected them.
* Generated xml code for the articles with unique ids starting from 700000.
* Xml code for first 10 birds: <https://github.com/Mahanth-Maha/indicwiki_birds/tree/main/curDir>

Dataset : [Birds Final](https://docs.google.com/spreadsheets/d/1H5m8lEYHytg9r5ao3jkkpSYcguiRcTVGTJtc3Rc52nI/edit?usp=sharing)

GitHub Link : <https://github.com/Mahanth-Maha/indicwiki_birds>