**Documentation on Unicode – Devanagari – English Transliteration**

Process: *Transliterating Unicode (Devanagari based) into English*

*Author: Ewan Nikhil Thomas*

Script to be used: *unicode\_english\_transliteration.py, spiders: devanagiri, english*

The NREGA assets data on Rajasthan contained Devanagari script based Unicode instead of containing English alphabets in the “panchayat\_name” column. Since this column was essential in uniquely identifying each row of observation, it was necessary that these Unicode values are transliterated into English. The process for the same is as listed below:

1. Two scrapy spiders namely, “devanagiri” and “english” were created to extract the Unicode value and its corresponding Devanagari string value. The spider “devanagiri” scrapes Unicode to Devanagari values from the [jrgraphix](https://jrgraphix.net/r/Unicode/0900-097F) webpage. This source was pointed out by Abhijit Parmar (RA to Prof. Ashwini Chhatre). The spider “english” scrapes Devanagari to English values from the ‘[easytyping’](https://www.easyhindityping.com/hindi-alphabet" \l "google_vignette) webpage which I found online.
2. The spiders created two .json files namely, “devanagiri” and “english” which will serve to make dictionaries for transliteration. Information from the two spiders were used create the csv file “data\external\devanagiri.csv” using the script “unicode\_english\_transliteration.py”.
3. Though transliteration details were available in the earlier mentioned webpage, there were some Devanagari special characters which required special attention. Transliteration of these characters were done manually using information from the ‘[Sanskrit](https://sanskrittestseries.com/wp-content/plugins/pramukhimeindic/pramukhime/help/pramukhindic-sanskrit.html)’ webpage.
4. Column values from the ‘devanagiri.csv’ file is used to create Unicode-Devanagari dictionary.

During LGD mapping of files, it was discovered that many states had multiple observations where panchayat names were in Unicode. Such instances were later addressed using the above process.