Create geonames to schema mapping

# 1 Workflow Geonames allCountries.txt to JSON file

Generate

1.1 Download geonames tab delimited files from http://download.geonames.org/export/dump/:

allCountries.txt

## 1.2 Escape double Quotes

## in allCountries.txt, alternatenames.txt

sed 's#"#\\"#g' allCountries.txt > allCountries\_esc.tsv

## 1.3 Concat header file to geonames file

cat gh.txt allCountries\_esc.tsv > allCountries\_esc\_gh.tsv

## 1.4 Use Karma and apply modelallcountries\_schema7.ttl

Create allcountries\_schema.json

## 1.5. Use Karma and apply model allcountries\_schema\_adm\_uri.ttl

Create allcountries\_schema\_adm\_uri.json

## 1.6 Use jq and sed to generate uri mapping file in blank separated quoted format

jq '"\(.uri)@\(.sameAs)"' allcountries\_schema\_adm\_uri.json | sed 's#@#\" \"#' > allcountries\_schema\_adm\_uri.csv

## 1.7 Replace temporary admin uri in allcountries\_schema.json with geonameid uris

cat allcountries\_schema\_adm\_uri.csv | while read line; do

neww=${line##\* };

oldw=${line%% \*};

sed "s# $oldw#$neww#g" allcountries\_schema.json > allcountries\_schema\_rep.json ;

rm allcountries\_schema.json;

mv allcountries\_schema\_rep.json allcountries\_schema.json

done

# 2 Workflow Geonames alternatenames.txt to JSON file

2.1 Download alternatenames tab delimited files from http://download.geonames.org/export/dump/:

alternatenames.txt

## 2.2 Escape double Quotes

## in alternatenames.txt, alternatenames\_esc.txt

sed 's#"#\\"#g' alternatenames.txt > alternatenames\_esc.tsv

2.3 Add header to alternatenames file

cat alternateNames\_header.txt alternateNames\_esc.txt > alternateNames\_karma.txt

2.4 Use karma and apply alternatenames1.tll model

to create alternatenames.json

# 3 Reduce alternatenames.json to allcountries\_schema.json

3.1 Create Json line files

jq '.[]' -c alternateNames\_samp.json > alternateNames\_samp\_line.json

jq '.[]' -c allcountries\_schema.json > allcountries\_schema\_line.json

3.2 Use Reducer of Karma to get Alternatenames in allcountries