

Assignment 1: Edit Distance

Provided at the end of this document are IPA transcriptions of 24 words from various Pashto dialects of 11 regions/cities along with their comparative words in Urdu/English . First two lines correspond to the longitude and latitude coordinates of the various cities.

Instructions

You need to carry out the following steps for completion of this task:

1. Write code for computing Levenshtein Distance (Edit distance)

Using the algorithms mentioned in slides, prepare code for computing the edit distance. Write your own code. Don't use ready-made libraries. You need to modify the edit distance metric such that it is able to report the distance as a similarity on a percentage scale.

2. Write a Calling code that is able to call the Distance metric multiple times (i.e for each pair of cities.

A table showing the average scores should be shown, where rows and columns mention names of each of the cities and a cell location corresponds to the average similarity measure (in percentage) between the two cities.

3. Bonus

A plot (similar to one shown in slide 24) showing the edit distance as a line between two cities. The thickness / opacity of the line can be adjusted based on the edit distance score.

Deliverable

Your deliverable should contain only 1 zip file. The name of the zip file should be your roll number. e.g. 20P-1234.zip. The zip file should contain the following files:

1. Your code (named as 20P-1234.py or 20P-1234.c or 20P-1234.cpp depending on which language you use)
2. Screenshot of generated table (named as 20P-1234.jpg)
3. Screenshot of generated map (if done. Named as 20P-1234-map.jpg)

Data

English	Quetta	Harnai	Bannu	Miranshah	Tanai Khost	Tirah
Longitude	66.9749	67.9374	70.6454	70.1730	69.8597	70.7517
Latitude	30.1798	30.1044	32.9909	32.9514	33.3585	33.8094
Pashto	Pašto	Pašto	Pāštē	Pāštē	Pāxtē	Pāxto
four	tsalor	tsaler	sālēr	tsālwēr	tsālēr	tsālwor
six	špaž	špož	špēž	špēž	špēg	špēg
how	tsənga	tsona	sərāng	tsərāng	tsərgē	tsərāng
we	muž	moš	miž	miž	mig	mu
my	zmā	mā eghē	ēmo	ēmo	ēmo	ēmo
your	stā	tāgha	ēto	ēto	ēto	ēto
boy	halək	waṛīz, čorī	wēṛkā	wēṛkai	wēṛkai	wēṛkai
girl	njiləi	čuwara	wēṛkyē	ǰəlkiyē	ǰəlkiyē	wēṛkyē
foot	pša	ghədəi	pšā	pšā	pšā	pxā

Sun who many little to drink egg yes I am I go tongue home it exists bear father	lmar tsok ḍēr zyāt ləḷ čšəl hagəi wo yəm dzəm žəba kor sta yiḷ plār	mērə čok tsaṭ ləḷ ghwətəl hoya wo ī drimī zbə kor sta yirz piyār	myērə stərgā sēk pirā zyot ləški čšəl yēya ēh yə sə žəbā kēr štā yiḷ plor	ghormə stərgā tsēk rəṭ zyot ləški čšəl yēya ēh yə tsə žəbā kēr štā yiḷ plor	myērə stərgā tsēk rəṭ zyot ləg tsəḵəl yēya ēh yəm tsəm žəbā kēr štā yig plor	myērə stərgā tsok ḍēr zyot ləg tsəḵəl woya ēh yəm tsəm jəba kolə štā yig plor
English	Urdu Karachi	Kandahar	Maidanshahr	Kabul	Peshawar	Marwat
Longitude Latitude Pashto four six how we my your boy girl foot Sun who many little to drink egg yes I am I go tongue home it exists bear father	67.0011 24.8607 Pašto char chay kesay hum mera tera ladka ladki pao suraj kon buhāt ziada thoda peena anda ha mai jana zuban ghar tera bhedia baap	65.7372 31.6289 Pašto tsalor špaḷ tsənga muḷ zmā stā halək njlai pša lmar tsok ḍēr zyāt ləḷ čšəl hagəi wo yəm dzəm žəba kor sta yiḷ plār	68.8407 34.3611 Paḵto tsalor špaḷ tsənga muḷ zmā stā halək jilai pḵa lmar tsok ḍēr zyāt ləḷ tsəḵəl hagəi wo yəm dzəm žəba kor sta yiḷ plār	69.2075 34.5553 Pəxto tsalor špag tsənga mung zəmə stā halək jinai pxa nmar tsok ḍēr zyāt ləg tskəl hagəi/hā wo yəm zəm žəba kor štā yig plār	71.5249 34.0151 Puxto salor špag singa mung zamā stā halək jinē xpa nwar sok ḍēr zyāt ləg skəl hagē/hā ao yəm zəm jəba kor štā yig plār	70.9012 32.6135 Pašto čālor špaḷ čəraṅga muḷ ēma ēta zərkayē jinkiyē pša lmar čok ḍēr zyāt ləḷ čšəl angəi aa yə jəm žəba kor sta iḷ plār