**Methodolgy for Phonetic matching**

**Challenge:**

One of the biggest challenge is matchingtwo set of strings which are not exactly same but are equivalent. For languages like Hindi where it is required to handle various pronunciations by people from various communities for the same keyword, especially in rural areas. This requires some sort of approximate matching algorithms.

**Methodolgy:**

There are many phonetic algorithms that are available for english language but no efficient algorithm available for indian languages,

The idea to implement an algorithm is to use soundex phonetic indexing algorithm which deals based on word pronounciation, soundex is actually implemented for english language, The same concept will be using to deal with hindi language,

* **Objective:**
* Comparison should be based on pronunciation

The following steps to be considered as:

* Group short and long vowel to a single code. e and ee is considered as equal.
* Group consonant families. ka, kha,ga,gha, nga becomes a single family. Same is the case of cha, ta, tha,pa.
* Group ra, Ra
* Group la,La, zha
* Group sa,Sa,sha

Grouping all together will be resulting in a mapping between hindi alphabets and phonetic codes can be called as character map.

The basic idea is to encode the hindi string based on the mapping generated by following above steps,

If two strings are having same phonetic codes then they are said to be equal else not equal.

**Algorithm Flow:**

* For each letter in the word except first letter, get the corresponding soundex digit from the character map, which is nothing but a table
* If the letter is not found in character map, the  soundex digit for that letter is 0
* Duplicate consecutive soundex codes are skipped. ie, effectively क्क will be considered as क
* Return soundex code padded to the required length (ie सLKES000) which is सन्तौष in hindi

E.g of how character map looks

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **अ** | | | **आ** | | | **इ** | | | **ई** | | | **उ** | | | **ऊ** | | | | **ऋ** | | | **ऌ** | | | **ऍ** | | | **ऎ** | | | **ए** | | | **ऐ** | | | **ऑ** | | | | **ऒ** | | | **ओ** | | | | **औ** | | |
| **A** | | | **A** | | | **B** | | | **B** | | | **C** | | | **C** | | | | **P** | | | **Q** | | | **0** | | | **D** | | | **D** | | | **D** | | | **E** | | | | **E** | | | **E** | | | | **E** | | |
| **क** | **ख** | **ग** | | **घ** | **ङ** | | **च** | **छ** | | **ज** | **झ** | | **ञ** | **ट** | | **ठ** | **ड** | **ढ** | | **ण** | **त** | | **थ** | **द** | **ध** | **न** | **ऩ** | | **प** | **फ** | | **ब** | **भ** | | **म** | **य** | | **र** | **ऱ** | **ल** | | **ळ** | **ऴ** | | **व** | **श** | **ष** | | **स** | **ह** |
| **F** | **F** | **F** | | **F** | **G** | | **H** | **H** | | **H** | **H** | | **G** | **I** | | **I** | **I** | **I** | | **J** | **K** | | **K** | **K** | **K** | **L** | **L** | | **M** | **M** | | **M** | **M** | | **N** | **O** | | **P** | **P** | **Q** | | **Q** | **Q** | | **R** | **S** | **S** | | **S** | **T** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ा** | **ि** | **ी** | **ु** | **ू** | **ृ** | **ॄ** | **ॅ** | **ॆ** | **े** | **ै** | **ॉ** | **ॊ** | **ो** | **ौ** |
| **A** | **B** | **B** | **C** | **C** | **P** | **P** | **E** | **D** | **D** | **D** | **D** | **E** | **E** | **E** |

The above english alphabets are the mappings for the corresponding hindi font