Match rating approach

From Wikipedia, the free encyclopedia
(Redirected from Match Rating Approach)

This article **provides insufficient context for those unfamiliar with the subject**. Please help improve the article with a good introductory style. *(October 2009)*

The **match rating approach** (MRA) is a phonetic algorithm developed by Western Airlines in 1977 for the indexation and comparison of homophonous names.[1]

The algorithm itself has a simple set of encoding rules but a more lengthy set of comparison rules. The main mechanism being the similarity comparison which calculates the number of unmatched characters by comparing the strings from left to right and then from right to left and removing identical characters. This value is subtracted from 6 and then compared to a minimum threshold. The minimum threshold is defined by table A and is dependent upon the length of the strings.

The encoded name is known (perhaps incorrectly) as a personal numeric identifier (PNI). The PNI codex can never contain more than 6 alpha only characters.

Match rating approach performs well with names containing the letter "y" unlike the original flavour of the NYSIIS algorithm. For example, the surnames "Smith" and "Smyth" are successfully matched.

MRA does not perform well with encoded names that differ in length by more than 2.

Contents [hide]

Encoding rules [edit]

1. Delete all vowels unless the vowel begins the word
2. Remove the second consonant of any double consonants present
3. Reduce codex to 6 letters by joining the first 3 and last 3 letters only

Comparison rules [edit]

In this section, the words "string(s)" and "name(s)" mean "encoded string(s)" and "encoded name(s)".

1. If the length difference between the encoded strings is 3 or greater, then no similarity comparison is done.
2. Obtain the minimum rating value by calculating the length sum of the encoded strings and using table A
3. Process the encoded strings from left to right and remove any identical characters found from both strings respectively.
4. Process the unmatched characters from right to left and remove any identical characters found from both names respectively.
5. Subtract the number of unmatched characters from 6 in the longer string. This is the similarity rating.
6. If the similarity rating equal to or greater than the minimum rating then the match is considered good.

Minimum threshold [edit]

The following table shows the mapping between the minimum rating and the string lengths.

Table A

Sum of Lengths	Minimum Rating

Wikipedia
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Languages
ไทย
Edit links

≤ 4	5
4 < sum ≤ 7	4
7 < sum ≤ 11	3
= 12	2

Match rating approach examples [edit]

The table below displays the output of the match rating approach algorithm for some common homophonous names.

Name	MRA Codex	Minimum Rating	Similarity Comparison Rating
Byrne	BYRN	4	5
Boern	BRN		
Smith	SMTH	3	5
Smyth	SMYTH		
Catherine	CTHRN	3	4
Kathryn	KTHRYN		

References [edit]

1. ^ Moore, G B.; Kuhns, J L.; Treffzs, J L.; Montgomery, C A. (Feb 1, 1977). *Accessing Individual Records from Personal Data Files Using Nonunique Identifiers*. US National Institute of Standards and Technology. p. 17. NIST SP - 500-2. Lay summary.

External links [edit]

- An Overview of The Issues Related to the use of Personal Identifiers, HSMD, Statistics Canada
- C# Implementation: http://sounditout.codeplex.com/

The Wikibook *Algorithm_implementation* has a page on the topic of: ***Match rating approach***

Categories: Phonetic algorithms