

Documentation for RB-AMIN . exe

Rikhil R. Bhavnani*
February 2009

Overview

This program identifies “fuzzy” or approximate matches for Indian names. Applications include identifying candidates that ran for office in two elections in India, as used in Bhavnani (2009). RB-AMIN . exe is useful since there is no standard transliteration for Indian names, because name parts are often transposed, and to take into account spelling errors.

Inputs

A source file which lists the names for which matches need to be identified, and a target file which lists the names amongst which matches need to be identified. Both files need to be in txt format. A record in each file should contain a number (which may be a record identifier, such as a candidate ID), followed by a name. The two fields must be delimited by a comma.

Output

Output.txt lists for each record in the source file matches in the target file and their match scores, provided match scores are greater than 90. The fields in the output file are as follows:

- Source record identifier
- Source name
- Target record identifier
- Target name
- Match score

Method

1. For each record in the target file, generate all name-part permutations, call this the modified target file (if there is a record for Raj Kapoor in the target file, a new record—with the same record identifier as Raj Kapoor—will be generated. The name for this record will be Kapoor Raj).
2. Compare every record in the source file and modified target file to generate match scores.
3. List all matches for which the match score is above 90.

Genesis

This program is a modified version of an AMIN (Approximate Matching of Indian Names) program developed by the Election Commission of India. The AMIN program has been modified to (i) take into account permutations of the placement of name parts, (ii) to compare two lists of names (rather than individual names), and (iii) to provide all matches and scores for which match scores are over 90. Programming was done in Visual Basic.

Citation

RB-AMIN . exe is a free contribution to the research community. Please cite it as follows: Bhavnani, Rikhil R. 2009. RB-AMIN . exe: A Tool to “Fuzzy” Match Indian Names. Stanford University, Stanford, CA. URL: <http://www.stanford.edu/~bhavnani>.

References

Bhavnani, Rikhil R. 2009. "Do Electoral Quotas 'Work' After They Are Withdrawn? Evidence from a Natural Experiment in India." *American Political Science Review* Forthcoming.

* PhD Candidate, Department of Political Science, Encina Hall West, Room 100, Stanford University, Stanford, CA 94305, rbhavnani@stanford.edu. Preliminary and incomplete. Comments appreciated. Please see citation details above.