

Extract ZIP Code

Description

This algorithm parses the address information provided and extracts ZIP codes from it. Currently it accepts ZIP codes which are in United States of America format i.e. either XXXXX (short form) or XXXXX-XXXX (long form).

Pros & Cons

This algorithm facilitates quick Spatial analysis by extracting ZIP codes from a given address, which can be further processed. Its only limitation is that currently it only supports parsing of USA ZIP Codes or countries which have USA based ZIP Code format.

Implementation Details

The algorithm works as follows,

1. Get the address for each row & begin parsing for zip codes in following manner,
 - a. Save all groups of digits along with their start position, end position & length of the group of digits.
 - b. Since the zip code, presumably, will be in the later portion of the address string, traverse the collected zip code candidates in the reverse fashion.
 - c. If there is a 5 digit group then,
 - i. Consider it as the primary zip code. If the user wants the ZIP code in truncated form then the algorithm will skip the checking for extension ZIP code.
 - ii. The extension of the ZIP code follows the primary zip code, so check if the previous group has length 4 and if so, then check if its distance from primary zip is less than or equal to 2. If yes, than consider this as the extension of the zip code.
 - iii. If there is no 4 digit group satisfying the above conditions then return null as the extension value.
 - iv. If there is no 5 digit group then return null for the primary zip code value. In this case, display a warning to the user that no ZIP code was found for this particular address string.

Usage Hints

The user has to provide 3 inputs; a file containing the addresses for which ZIP code parsing is required, whether to truncate the parsed ZIP code or not and name of the address column. If the plugin was unable to find any ZIP code then it will print a warning message and set the ZIP code to empty string. The data for ZIP codes can be in either short form i.e. XXXXX or long form i.e. XXXXX-XXXX. It will also accept ZIP code information in the following format,

XXXXX<Any Character(s) of Max Length = 2>XXXX.

The output of this algorithm will be the original input table with 1 column added containing the parsed ZIP code.

Links

- [Source Code](#)

Acknowledgments

The ZIP code extraction algorithm was authored, implemented, integrated and documented by Chintan Tank.

See Also



The license could not be verified: License Certificate has expired! [Generate a Free license now.](#)