

CS412 – Mini MP 2

Name: Guangyu Zhou
Netid: gzhou6

1. Unmarked

The workflow diagram shows a sequence of steps: ReadSource, MatchLastName, Select Suspects, and Suspects. A 'Lkp_LastName' step is connected to 'MatchLastName', and a 'Discard' step is connected to 'Select Suspects'.

Fuzzy string search

Step name: MatchLastName

General Fields

Lookup stream (source):

Lookup step: Lkp_LastName

Lookup field: match_lastname

Main stream:

Main stream field: src_lastname

Settings

Algorithm: Jaro Winkler

Case sensitive: ☐

Get closer value: ☐

Minimal value: 0.8

Maximal value: 0.99

Values separator: ,

Help OK Cancel

Execution Results

Execution History / Logging / Step Metrics / Performance Graph / Metrics / Preview data

First rows Last rows Off

#	src_custid	src_lastname	src_email	match
1	2	JOHNSON	PATRICIA.JOHNSON@sakilacustomer.org	JONES,JOHNSTON
2	3	WILLIAMS	LINDA.WILLIAMS@sakilacustomer.org	WILLIS,WILLIAMSON
3	4	JONES	BARBARA.JONES@sakilacustomer.org	JOHNSTON,JOHNSTON
4	5	BROWN	ELIZABETH.BROWN@sakilacustomer.org	BONE,BURTON,BOWENS,BOWMAN,BROWNLEE
5	6	DAVIS	JENNIFER.DAVIS@sakilacustomer.org	DANIELS,DAVIDSON
6	7	MILLER	MARIA.MILLER@sakilacustomer.org	MILES,MILLS,MILNER,MOELLER,MILLARD
7	8	WILSON	SUSAN.WILSON@sakilacustomer.org	WILES,WILLIS,LAWSON,OLSON
8	9	MOORE	MARGARET.MOORE@sakilacustomer.org	MORRELL,MORALES
9	11	ANDERSON	LISA.ANDERSON@sakilacustomer.org	GUNDERSON2012,ANDREWS,PEARSON,ANDREW,SANDERS,LARSON,PATTERSON,HEP
10	12	THOMAS	NANCYTHOMAS@sakilacustomer.org	THOMPSON

2. Marked

The workflow diagram is identical to the one in the 'Unmarked' section, showing steps: ReadSource, MatchLastName, Select Suspects, and Suspects, with 'Lkp_LastName' and 'Discard' steps.

Fuzzy string search

Step name: MatchLastName

General Fields

Lookup stream (source):

Lookup step: Lkp_LastName

Lookup field: match_lastname

Main stream:

Main stream field: src_lastname

Settings

Algorithm: Jaro Winkler

Case sensitive: ☐

Get closer value: ☒

Minimal value: 0.8

Maximal value: 0.99

Values separator: ,

Help OK Cancel

Execution Results

Execution History / Logging / Step Metrics / Performance Graph / Metrics / Preview data

First rows Last rows Off

#	src_custid	src_lastname	src_email	match	score
1	2	JOHNSON	PATRICIA.JOHNSON@sakilacustomer.org	JOHNSTON	1
2	3	WILLIAMS	LINDA.WILLIAMS@sakilacustomer.org	WILLIAMSON	1
3	4	JONES	BARBARA.JONES@sakilacustomer.org	JOHNSTON	0.8
4	5	BROWN	ELIZABETH.BROWN@sakilacustomer.org	BROWNLEE	0.9
5	6	DAVIS	JENNIFER.DAVIS@sakilacustomer.org	DAVIDSON	0.9
6	7	MILLER	MARIA.MILLER@sakilacustomer.org	MILNER	0.9
7	8	WILSON	SUSAN.WILSON@sakilacustomer.org	WILES	0.9
8	9	MOORE	MARGARET.MOORE@sakilacustomer.org	MORRELL	0.8
9	11	ANDERSON	LISA.ANDERSON@sakilacustomer.org	ANDREWS	0.9
10	12	THOMAS	NANCYTHOMAS@sakilacustomer.org	THOMPSON	0.9