

Privacy Preserving Record Linkage Mihir Kumar

Mentor: Vincent Bindschaedler

Motivation and Objective

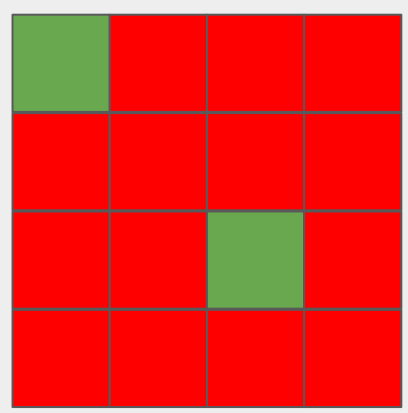
Identifying records that refer to the same individual from multiple databases without revealing the identity of the person nor the contents of the database

Overview

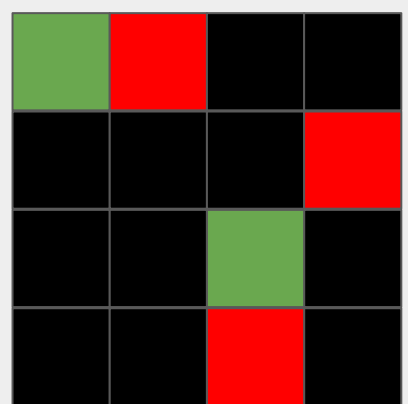


Blocking

No blocking



With blocking



Field Comparison

First Name Last Name

John Smith

Jon Smyth

Similarity Function

0.75 0.8

Field Comparison Vector

Record Pair Comparison

First Name Last Name

John Smith

Jon Smyth

Similarity Function

0.75 0.8

Field Comparison Vector

Record Pair Similarity "Score": 0.79

Record Pair Classification

ABC Records	XYZ Records	RPSS
John Smith	Jon Smyth	+7
John Smith	Taylor Swift	+3
Lucy Ball	Mario Kart	0
Lucy Ball	Taylor Swift	0

Record Pair Classification

+7	Match!
+3	No Match
0	No Match
0	No Match

Binary Field Comparison

John Smith Urbana

Jon Smyth Urbana

SHA

xy9l br3f xt0uv

nw2 vwer xt0uv

Equal?

0 0 1

Field Comparison Vector

Approximate Field Comparison

J JO OH HN N

1 0 1 0 0 1 1 1 0 1

J JO ON N

1 0 1 0 1 1 0 1 1 1

Dice coefficient = $(2 * 5) / 13 = 0.77$

Fellegi-Sunter (FS)

First Name Last Name

John Smith

Jon Smyth

0 0

-0.68 -1

+ +

=

-1.68

Winkler's Modification to FS

First Name Last Name

John Smith

Jon Smyth

0 0

-0.68 -0.68

0.73 1.36

+ +

=

2.09

