Technical Integration Guide

Before you get started:

* Html must be in current location of notebook
* Html must be run in firefox and launched by the Jupyter notebook, MatchNMerge.ipynb. HTML will run locally (e.g. <file:///C:/Users/rlim/fuzzy/fuzzyjoinerv1.html>)
* Write to status.txt whatever you want and whenever you want so that it will be displayed as status notes in the UI (reads file periodically)

Step 1:

* Load data set 1 and data set 2 (must be csv files)
* Must load csv file to current notebook folder
* Step 1 will generate filesready.txt (this is a comma-delimited list of file names – e.g. dataset1.csv, dataset2.csv)

Step 2:

* Click on “Join Columns”
* This should run Stephanie and Jordan’s code
  + Code should be waiting for filesready.txt
  + This sample py code waits for the existence of a file:
    - #TODO: you wait for a file called "filesReady.txt".
    - # This file will contain two entries, separated by a comma. They correspond to the dataset files.
    - # The files will be located at the same location of this notebook
    - # some code I pulled from the internet to wait for a file before proceeding
    - import os.path
    - import time
    - file\_path = "filesReady.txt"
    - print('going in')
    - while not os.path.exists(file\_path):
    - time.sleep(5);
    - if os.path.isfile(file\_path):
    - continue;
    - else:
    - time.sleep(1);
  + Process code
    - When you are done, you should generate a file called columnMapping.csv.
      * First column: column names of dataset 1
      * Second column: mapped column names of dataset 2
      * Third column: Relative accuracy

Step 3:

* User can go on to step 3 when it sees that columnMapping.csv has arrived
* User clicks on “Display Column Mappings” button
* User is presented table grid in UI
  + Final column is added by UI
  + User can leave “Keep Match” or change to modify the mapping
  + Note: The drop down of columns that a column mapping can be changed to are the columns from data set 2
  + (TO BE ADDED) User can set “Do Not Match” if he/she does not want to match
* When user is ready to finalize, click on “Finalize Column Mappings” button
* This generated columnMapping\_user.csv file
  + This file is essentially columnMapping.csv with an added column ‘Change Column Mapping’
    - If “Keep Match” set, use mapping from column 2
    - If “Do Not Match” set, don’t match
    - If something else, use that column for matching

Step 4:

* Rob’s code waits for columnMapping\_user.csv
* Does join row code
* UI waits for file results.csv. When it sees that, it displays a link to the user will be allowed to download results.csv