Homework 10 (Due at 11:59pm on 2015/04/03)

## Instructions:

In this homework, you will need to use **FRIL** (<http://fril.sourceforge.net>) and try different metrics to link the records between the two given restaurant datasets: **d2.csv** and **d1.csv**.

For evaluation of your proposed metrics, the set of correct links: **groundtruth.csv** is also given.

You need to correctly match as many records as possible. However, you are **NOT** allowed to change any data of the input files (20 points will be deducted if you do otherwise).

We will grade your submission based on your model’s F1 score[1] on the dataset. The grading policy is as follows.

* We will calculate average the F1 score (μ) of all submissions and its standard variance (σ).
* The submission with no less than the average F1 score will get full credits (60 points)
* The submissions with F1 score lower than the average get 10 points deducted for every standard variance lower. Ex: if your score (v) is below μ, your score is 60 - [(μ-v)/σ + 1]\*10

Reference:

[1] F score: <http://en.wikipedia.org/wiki/Precision_and_recall>

## Submission guideline:

* **Hw10.pdf**: a report containing description of the metrics that you used for each field and your F1 score. Briefly explain why you choose the metrics and the way you use to calculate the F1 score. (20 points)
* **results.csv**: the output results from FRIL which achieved the best F1 score performance. The results file should show the matched records in pairs using their line numbers. Each matched pair in its own line. For example,

13, 27

34, 49

…

it means the 13th line of d2 is matched with the 27th line in d1. Please make sure that your file is in this format.

* **configuration.xml**: the FRIL XML configuration file ( 5 points)
* **Your code** that processes the output file of FRIL and converts it into the results.csv

Zip all your files into a single file that should be named **hw10\_[firstname]\_[lastname].zip** and submit on Blackboard.