# Homework 6: Data Cleaning

Due Date: Sunday, 2015/10/28 @11:59pm on Blackboard.

## Summary

In this homework, you will perform data cleaning on the data you have extracted in a previous homework (e.g. for wrappers) or for your project.

## Task 1 – Data Cleaning

Identify at least **3 fields** that requires cleaning in your data.

> E.g. Splitting the name field into first name and last name.

Perform the cleaning using any tools taught in class. One possible tool you can use is OpenRefine (http://openrefine.org).

## Task 2 – Questions

1. For each field you chose, describe what it is and what was the operation performed to clean it.
2. Include an example of a record from each chosen field before and after cleaning in your report.

## Submission Instructions

You must submit the following files (totally 2 files) in a single .zip archive named `Firstname_Lastname_hw6.zip` and submit it via Blackboard:

- The PDF file `Firstname_Lastname_hw6.pdf` containing the answers to Task 2. **(30 points)**
- The data file `Firstname_Lastname_hw6_data` containing the data you have cleaned with an appropriate extension like .xml or .json. **(70 points)**

## Ground Rules

This homework must be done individually. You can ask others for help with the tools, but the submitted homework needs to be your own work.