Deliverable 1: **Manipulate Street Address**

Assigned Date: **Wed/Thu, 2015-01-21/22** (week 02)

Due Date: **Wed, 2015-02-04, 08:30 AM** (week 04)

Grade Value: **20%** of Final Grade

# Important

1. *Plagiarism is a serious offense and will not be tolerated! Plagiarized assignments will receive a grade of 0 (zero) as per college policy. In addition, professionalism/participation grades will be reduced to 50%. Please refer to the official college policy on academic misconduct which can be found here:* [*http://www.niagaracollege.ca/Content/LinkClick.aspx?fileticket=i-eJpagoNDQ%3d&tabid=1001*](http://www.niagaracollege.ca/Content/LinkClick.aspx?fileticket=i-eJpagoNDQ%3d&tabid=1001)*.*
2. *Save your work periodically.*
3. *The assignment will be evaluated in terms of both functionality and industrial standard coding style.*

# Background:

This assignment provides a basic introduction to Python scripting, especially in string manipulation, which is very fundamental and important in preparing raw data for various GIS geoprocessing tasks.

# Purpose:

To separate a whole piece of mailing address into different address components, such as street number, street name, street suffix type, street direction, city, province, and postal code.

# Procedure:

1. Create a subfolder named \LastNameInitialGisc9307D1 in C:\temp (create it if necessary) and copy the \assignments\d1ManipulateAddress\d1RawData subfolder within the newly created subfolder. Noticing that the \assignments\d1ManipulateAddress\ is the path and the d1RawData is the subfolder name.
2. Carefully investigate the two text files that are contained within the \d1RawData subfolder (hint: it would get better view by loading the text file into Excel Workbook):
   1. One is the raw data, \d1RawData\d1RawListOfFarms.txt: it contains **a raw list** of 91 mailing addresses of farms that are located in the Niagara Peninsula; it has 2-column and is tab-delimited. A small set of records is shown in Table 1. The original addresses are derived from <http://www.harvestontario.com/>.
   2. The other is the example of the expected result after the script is run, \d1RawData\d1ResultListOfFarms.txt: it is **a final processed list** of 91 mailing addresses of farms, containing 9-column of different address components. A small set of records is shown in Table 2. This text file serves as an example and your own processed result will be saved in another folder.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1. Original Raw Data** | | 🡺🡺🡺 | **Table 2. Data that Have Been Processed After Running the Scripts** | | | | | | | | |
| **ID** | **Address** | **ID** | **Address** | **StreetNum** | **StreetName** | **SufType** | **Dir** | **City** | **Province** | **PostalCode** |
| 5 | 5172 Green Lane Rd, Lincoln, ON L0R 1B3 | 5 | 5172 Green Lane Rd, Lincoln, ON L0R 1B3 | 5172 | Green Lane | Rd |  | Lincoln | ON | L0R 1B3 |
| 6 | 500 Glenridge Ave, St. Catharines, ON L2S 3A1 | 6 | 500 Glenridge Ave, St. Catharines, ON L2S 3A1 | 500 | Glenridge | Ave |  | St. Catharines | ON | L2S 3A1 |
| 7 | 471 Foss Rd, Pelham, ON L0S 1C0 | 7 | 471 Foss Rd, Pelham, ON L0S 1C0 | 471 | Foss | Rd |  | Pelham | ON | L0S 1C0 |
| 9 | 3836 Main St, Lincoln, ON L0R 1S0 | 9 | 3836 Main St, Lincoln, ON L0R 1S0 | 3836 | Main | St |  | Lincoln | ON | L0R 1S0 |
| 13 | 1167 Lakeshore Rd W, St. Catharines, ON L2R 6P9 | 13 | 1167 Lakeshore Rd W, St. Catharines, ON L2R 6P9 | 1167 | Lakeshore | Rd | W | St. Catharines | ON | L2R 6P9 |
| …… | |  | …… | | | | | | | | |

1. Use the contents of the textbook (chapters 1, 4, 7.6, and other relevant online resources) to write a piece of Python script, named as LastNameInitialGisc9307D1.py in C:\temp\LastNameInitialGisc9307D1 subfolder, to:
   1. Create a subfolder d1ProcData within \LastNameInitialGisc9307D1 subfolder (using python script). Delete it if it exists before create it. The \d1ProcData subfolder is used to save your own processed results.
   2. Separate the originally provided mailing addresses of 2-column structure into that of 9-column one.
   3. Write the processed result as a plain text file d1ResultListOfFarms.txt in \d1ProcData subfolder. The expected result has 9 columns which are tab-delimited.

# Special Notes:

* 1. Make sure your codes will be working properly as your submission folder \LastNameInitialGisc9307D1 subfolder is copied to the instructor’s c:\temp subfolder.
  2. Avoid the hard coding. For example, as more streets suffixes (such as Cres, Cir, Trl, and Crt) are added in the raw text file, the codes should be run smoothly with minimum modification.
  3. The processed result should be clean. For example, the result should contain neither extra empty lines nor extra spaces;
  4. Carefully examine the d1ResultListOfFarms.txt file to ensure the desired result is expected.

# Deliverable:

Zip the whole subfolder \LastNameInitialGisc9307D1 and name it as LastNameInitialGisc9307D1.7z using 7-Zip application (do not use WinZip, WinRar, or other compression applications). Submit the .7z file to your instructor by email at [jjiang@niagaracollege.ca](mailto:jjiang@niagaracollege.ca).

# Marking Scheme:

* 1. 10% Email etiquette and formal transmittal letter;
  2. 10% File and folder structure;
  3. 80% Functionality and professional coding style.