Debtor Identification

Basics:

The library used in this project is R.Net which is an interoperability library to access R from .net languages. It can be installed from NuGet packages under R.Net.Community.

- Commands to initialize R engine:

```
var envPath = Environment.GetEnvironmentVariable("PATH");
var rBinPath = @"C:\Program Files\R\R-3.3.0\bin\i386";
Environment.SetEnvironmentVariable("PATH", envPath +
System.IO.Path.PathSeparator + rBinPath);
engine = REngine.GetInstance();
engine.Initialize();
```

- Getting the directory of the bin folder:

```
string cur_dir = System.IO.Directory.GetCurrentDirectory();
cur_dir = cur_dir.Replace("\\", "/");
```

- Load R source file:

```
engine.Evaluate(string.Format("source('{0}/RData.R')", cur_dir));
```

- Load Dictionary:

```
engine.Evaluate(string.Format("Dictionary <-
read.table('{0}/Dictionary.txt', stringsAsFactors = FALSE)",
cur_dir));
```

Functions:

1- Single Algorithm functions:
All the functions using a single algorithm have the same following structure:

```
Results_list <- function(Err_data, Dictionary, Method)
```

- Err_data: A string or vector of strings containing the erroneous data
- Dictionary: A list of three vectors containing individual names, transliterated names and the frequency of each name.
- Method: can accept one of the following values (“dl”, “lcs”, “1gram”, “2gram”, “3gram”, “cosine”, “jaccard”, “jaro”, “jw”)
- Results_list: A list of two vectors the first contains the matches found and the second contains the confidence measure.

- R functions:

- Match_Single_Name: returns a list of 5 matches and their confidence.
- Match_Single_Name_Soundex: returns a vector of n matches
- Match_Full_Name: returns a list of 5 matches for each word in the name and confidence, so the length of the vectors equals 5 multiplied by the number of words in the name.
- Match_Full_Name_Soundex: returns a list of matches for each word in the name.
- Match_Full_Name_top_25: returns the top 25 suggested combinations of the corrected name and their frequencies.

- Process_List_Method: the input is a vector of full names, the output is a list of 25 suggestions for each full name. The length of the vectors equals 25 multiplied by the number of names.

- Example:

```
engine.Evaluate("results_list <- Match_Full_Name(Err_data,
Dictionary, Method = 'jw')");
List <string> matches = new List <string>
(engine.Evaluate("results_list [[1]]").AsCharacter());
List <double> confidence = new List
<double>(engine.Evaluate("results_list [[2]]").AsNumeric());
```

2- Functions using a combination of 7 algorithms:
All functions the combination have the following format:

```
Results_list <- function(Err_data, Dictionary)
```

The same explanations apply.

- R functions:

- Match_Single_Name_Comb: returns a list of 5 matches and their confidence.
- Match_Full_Name_Comb: returns a list of 5 matches for each word in the name and confidence, so the length of the vectors equals 5 multiplied by the number of words in the name.
- Match_Full_Name_Comb_top_25: returns the top 25 suggested combinations of the corrected name and their frequencies.
- Process_List: the input is a vector of full names, the output is a list of 25 suggestions for each full name. The length of the vectors equals 25 multiplied by the number of names.

- Example:

```
 engine.Evaluate("results_list <- Match_Full_Name_Comb(Err_data,
Dictionary)");
 List <string> matches = new List <string>
(engine.Evaluate("results_list [[1]]").AsCharacter());
List <double> confidence = new List
<double>(engine.Evaluate("results_list [[2]]").AsNumeric());
```

3- Update Dictionary:

```
Dictionary <- Update_Dictionary(Char_Vec, Dictionary)
```

- Char_vec: a vector of full names from a list of confirmed debtors.

```
engine.Evaluate("Dictionary <- Update_Dictionary(Char_Vec,
Dictionary)");
engine.Evaluate(string.Format("write.table(Dictionary, file =
'{0}/Dictionary.txt')", cur_dir))
```