BEW/EEEP Data Extraction

1. Introduction:

BEW/EEEP are former programmes of BEC from 2009 to 2013.

1. Goal:

The purpose of the project is to extract data and review for Brian O’ Mahony (Energy Efficiency Finance) to evaluate the cost credits and measures undertaken by BEW/EEEP over the last number of years. He will use extracted data for informing decisions on the direction of future users of investment and understanding where the market in relation to different upgrade measures at present.

1. Brief Analysis:

The application is developed to get data of all BEW/EEEP files since 2009 to 2013. Tabs in each file need to be extracted are Admin, Summary (in Batch files), Technologies and BE Workplaces main workbook

In Admin tab, ‘Reference No.’, ‘Cat.’, ‘Submitted By’, ‘Project Title’, ‘County’, ‘Approved Funding’ are columns need to be extracted while ‘Reference’, ‘Applicant’, ‘Description’ are extracted columns for Summary tab. For both Technologies and BE Workplaces main workbook need many columns which are listed.

1. Resources used:

* Programming Language: Python 3.7
* Software:
* Python 3.7
* Excel
* Co-operators:
* Database: Liam Costello
* Customer: Brian O’ Mahony

1. Approach:

Here is a flow chart representing the approach of the application

1. Summary of outcome:

* Have successfully extracted data from approximately 35 files over 4 years since 2009 to 2013 into files which are Admin.xlsx, Summary.xlsx, Technologies.xlsx, Workplaces.xlsx and SEEEP Project and technology monitoring summary january 2010 DM.xlsx. Each file is extracted by taking the column names so the number of data points won’t change.

1. Technical notes:
2. Notes:

Manual changes:

* Manually saved all files in SEEEP without changing anything. Doing this will somehow make the script run without bugs.
* Move folder BEC 2013 into SEEEP folder and add ‘EE’ in its name as this folder contains required data
* Manually copied and saved SEEEP Project and technology monitoring summary january 2010 DM.xlsx as this file does not require any changes

Automatic changes:

* Add a new column for each file which is ‘Year’

1. Assumptions:

* All files in a year have the same format. However, there are some exceptions.
* All tabs in a year have the same format. However, there are some exceptions.
* There will be slight changes between years in the way the name of headers, the index changes. However, the order of columns should not be changed.
* If there is any different between headers of each tab, then that problem must be dealt manually as the script doesn’t have function for that
* The folder name of BEW and SEEEP data is SEEEP
* The folder name of BEW in year must be ‘BEW <year>’
* The folder name of SEEEP contains ‘EE’
* The outputs will be written into “.xlsx” files

1. How to install Python 3.7:

* Download link: <https://www.python.org/downloads/release/python-372/>
* Choose [Windows x86-64 executable installer](https://www.python.org/ftp/python/3.7.2/python-3.7.2-amd64.exe) if you are using windows
* Instruction: <https://realpython.com/installing-python/#step-1-download-the-python-3-installer> .Under Window section and start with step 1 then step 2.
* After installing., launch CMD -> type python

1. Install Pycharm (or any editor):

* Download link: <https://www.jetbrains.com/pycharm/download/#section=windows>
* Choose Community version
* Instruction: <https://beginnersbook.com/2018/01/python-install-pycharm-windows-mac-linux/>

1. How to run the application:

* Before running, please make sure that all modules are imported (Go to File -> Settings -> Project: [name of project] -> Project Interpreter -> “+” symbol -> search for missing modules which are underlined as red)
* Change the path that contains the folders BEC [year]. For example:

path = os.path.join(**'C:/Users/pphuc/Desktop/Docs/Current Using Docs/'**)

* Then Press Run on the top right of editor