Design and development of a data-extraction tool for digitizing and compiling building energy and emissions data

EXECUTIVE SUMMARY

Prepared by:

Ayush Suhane, UBC Sustainability Scholar, 2021

Prepared for:

Fred Tewfik, Manager, Building Approvals, City of Richmond

Nicholas Heap, Sustainability Project Manager, City of Richmond

August 2021

DISCLAIMER

This report was produced as part of the UBC Sustainability Scholars Program, a partnership between the University of British Columbia and various local governments and organisations in support of providing graduate students with opportunities to do applied research on projects that advance sustainability across the region.

This project was conducted under the mentorship of City of Richmond. The opinions and recommendations in this report and any errors are those of the author and do not necessarily reflect the views of the City of Richmond; or the University of British Columbia.

ACKNOWLEDGEMENT

The author acknowledges that the work for this project took place on the unceded ancestral lands of the xwməθkwəy̓əm (Musqueam) Nations.

The author would like to thank the mentors Fred Tewfik and Nicholas Heap from the City of Richmond for their constant support and encouragement during the project. The author would also like to thank Karen Taylor from the UBC Sustainability Initiative for the frequent project check-in sessions and organizing policy and project related seminars.

Cover photo by Matias Santana on Unsplash

Introduction

The Province of British Columbia (*BC*) introduced BC Energy Step Code ("*Step Code*") in 2017 as an optional compliance path to provide incremental guidelines promoting energy efficient building construction by establishing a series of measurable performance requirements with an eventual goal of reaching net-zero energy performance by 2032 (Energy Step Code, 2018). Several local governments in BC have adopted Step Code and require builders to demonstrate that the new buildings meet the Step Code requirements. To facilitate the review process of building permit applications and assess the overall performance of municipalities, it is useful to track building energy emissions in local as well as on a provincial level.

The City of Richmond, among other municipalities in BC, accepts extensive building energy modelling reports that include valuable details about the design, energy performance, and operational emissions of the building. However, given the volume of the development and building permit applications and the extensiveness of the data, manual extraction and compilation of the data are cumbersome. Moreover, despite continued efforts to digitize its intake and review process, the City still receives a significant number of paper reports, which adds difficulty of extracting and archiving the relevant data.

This project explores the possibility of designing and developing a prototype for a data-mining tool to extract desirable fields from energy modelling reports and store them in a database. To examine the feasibility of the prototype, the focus of this report is limited to a standard building energy reports of single detached houses in a portable PDF format.

Background

For building permit applications of single detached houses, the City of Richmond accepts several documents during different stages of the construction process i.e., pre-construction, mid-construction, and house completion, to review the building performance in comparison to the BC Energy step code requirements. Among these documents, three energy modelling reports i.e., HOT2000 report for reference house, HOT2000 report for proposed/as-built house both generated from an energy modelling software (HOT2000 developed by Natural Resources Canada), and a Part-9 Compliance report ("*Compliance*") are particularly important as they have all the necessary information related to the design and operational emissions of the building. Some of the relevant entries in these reports are annual consumption of electricity, and natural gas, green house gas emissions and other performance metrics such as Mechanical Energy Use Intensity (MEUI) and Thermal Energy Demand Intensity (TEDI) which are extremely useful quantities to satisfy Step Code requirements. The structure of the HOT2000 reports and Compliance reports is based on the structured templates also defined in the Step Code Instruction Manual (Compliance, 2021).

In general, manual extraction from the energy reports is a daunting task. The HOT2000 modelling reports can be as long as 30 pages in length, while Compliance reports are 3-page long summary reports. These structured reports can be used for automatic extraction of data using data-mining techniques.

Methodology

With the help of the mentor's contacts and consultation with several building approval offices in BC, a qualitative assessment is performed to understand the state-of-the art data-collection practices. A list of approximately 45 parameters related to the building energy design, emissions were tabulated in consultation with the mentors. A Windows-based application ("CORFieldExtractor") is developed in python environment to extract the tabulated parameters from the HOT2000 reports and Compliance reports of single-detached houses. The application is verified manually for all the identified fields. A step-by-step instruction manual is created to facilitate easy access for new users or municipalities. A more detailed technical manual is created highlighting the limitations, and scope for further developments in the application.

For the implementation of the application, a number of open-source python libraries such as PySimpleGUI (PySimpleGUI, n.d.) for graphical interface, PDFPlumber (PDFPlumber, n.d.), Camelot (Camelot, n.d.) for text extraction from pdf documents, and Pyinstaller (Pyinstaller, n.d.) to create an executable application were used. It should be noted that the prototype doesnot support data-extraction from images/scanned documents in the current version.

Results

Our preliminary scan of data-collection practices and discussions with other municipalities revealed significant interest in the development of a data-mining tool to assimilate the relevant

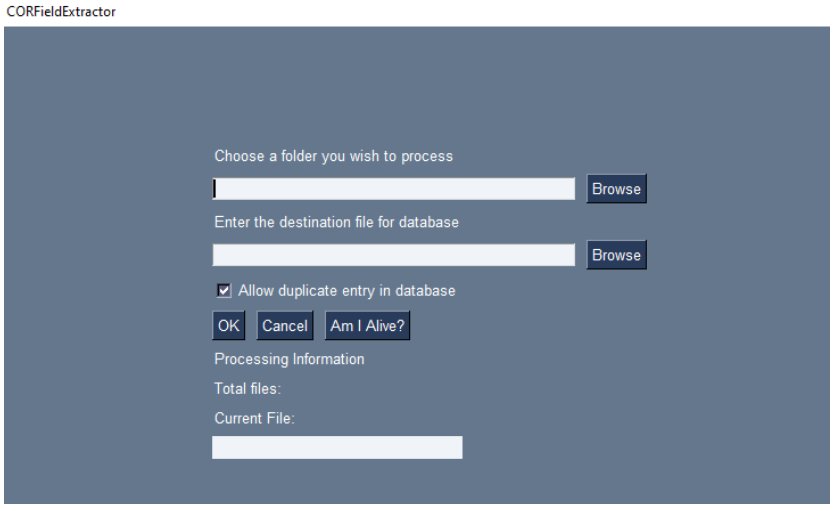

Figure 1. Graphical User Interface for the data-mining application

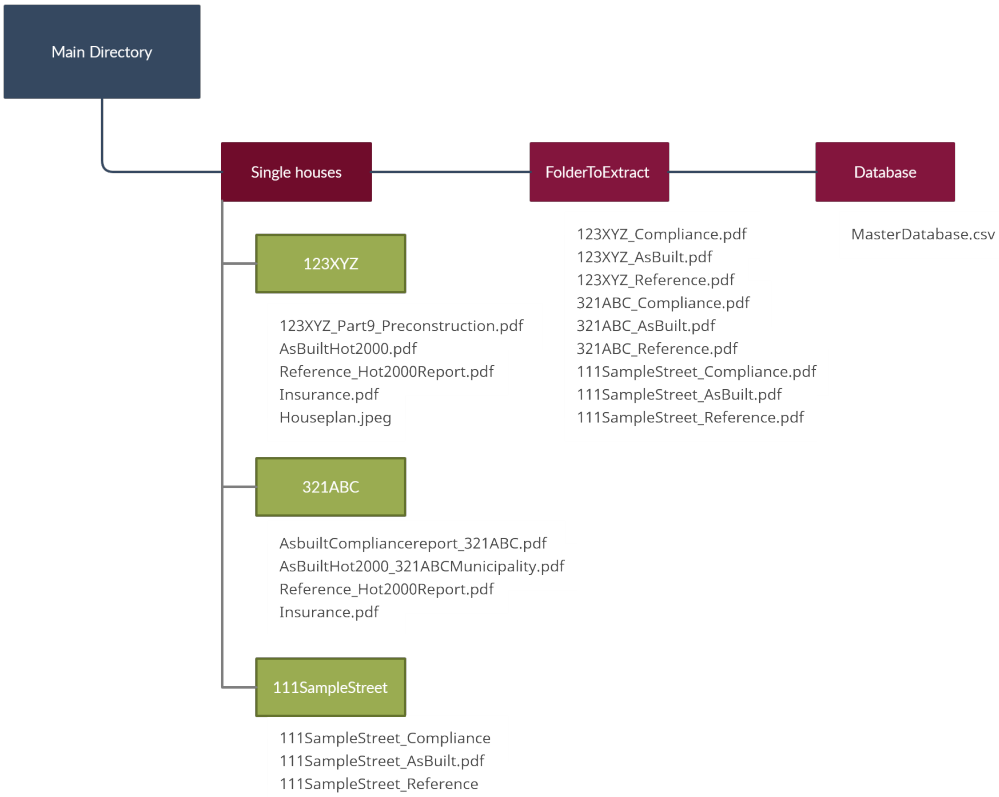

Figure 2. Folder structure for data-extraction

energy and performance metrics from several energy modelling reports. A particular roadblock in the development of the data-mining tool was to standardize the energy modelling reports received by the building approvals office. While there are several ways to create the document, we recommend building approvals office to accept the digital reports generated from HOT2000 software and Compliance report generated either from the excel tool or PDF form available on the BC energy website(Energy Step Code, 2018).

A screen-capture of the Graphical User Interface ("GUI") is shown in Fig. 1. The interface at the top lets the user select a source folder with all the files for extraction, and destination database file (".csv") in comma-separated values extension. The checkbox below these buttons enables the functionality to allow duplicate entries in the database. Submitting the form with "OK" starts the extraction process. "Am I Alive?" button indicates the responsiveness of the tool. Other entries at the bottom including a progress meter provides real-time information about the extraction process i.e., total number of files to be extracted, file name and percentage to completion. A file structure, as shown in Fig. 2, is also recommended to organize the information in the database. The application recognizes different types of files based on a naming convention, where the file name

should always begin with a house address followed by the file type i.e., “Compliance”, “Reference”, and “AsBuilt” separated by an underscore as shown in Fig. 2.

After the extraction is complete, all the relevant values are stored in the database (CSV file) for further inspection. Fig. 3 shows a sample output for three houses shown in Fig. 2. The top row includes abbreviations to different fields in the energy reports. The second row lists the source of the fields i.e. “C” refers to Compliance report, “H2K-AB” refers to HOT2000 report for as-built house, and “H2K-Ref” to HOT2000 report for reference house. More details on the abbreviations are provided in the technical report.

The efficiency of the tool is observed to be close to 90%. Several tokens such as “BLANK” for no entry in the document, “NA” for unavailability of the document or scanned document and “ERROR” for the inability of the tool to reliably extract data from energy reports are added to the database. Manual intervention is recommended for these tokens in the database.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	FILEKEY	Address	PostalCod	Permit	Version	Municipal	CEA	Builder	Nfile	Step	HeatedAr	AB_Airtigl	AB_REC	REF_REC	TEDI
2	-	C	C	C	C	C	C	C	C	C	C	C	C	C	C
3	111Sampl	[redacted]	[redacted]	BLANK	11.7	Richmond	[redacted]	[redacted]	[redacted]	1	NA	2.8	141	146	74
4	123XYZ	[redacted]	[redacted]	BLANK	11.8	Richmond	[redacted]	[redacted]	[redacted]	1	NA	3.8	120	121	72
5	321ABC	[redacted]	[redacted]	[redacted]	11.8	Richmond	[redacted]	[redacted]	[redacted]	1	NA	3.32	94	100	61

Figure 3. Screenshot of sample output in destination database file.

Summary

In this project, we undertook the development of a prototype to automate the data extraction from building energy reports submitted by the builders to the building approvals office. The tool parses the reports and extracts the information reasonably well with an accuracy of 90%. Several sources of errors were identified and highlighted in the database, where manual intervention can be useful. One of the major challenges in automating the data-extraction process is frequent modifications to the structured templates of the energy reports by the Government of British Columbia. The scripts for the data-extraction are designed to accommodate such changes with minimal modifications in the code. A complete package with user documentation (instruction manual, technical report with detailed comments about the code and references) is created to enable easy adoption for new users and to allow advanced users to add other functionalities to the application.

Next Steps

With the development of a prototype, several interesting avenues can be explored:

- The application along with user manual should be distributed to other municipalities to verify the data-extraction.
- A successful verification would enable an assimilation of energy and emissions data locally as well as on a provincial level.
- Additional functionalities such as extension to townhouses, as well as the ability to read scanned documents should be added to the application.
- The database should be used to generate insights about the energy emissions and overall progress of the government to reach the goal of net-zero ready buildings.

References

Energy Step Code. Energy Step Code | Government of British Columbia. (2018, March 3). https://energystepcode.ca/.

Compliance. Compliance tools for Part 9 buildings: Energy Step Code. Energy Step Code | Government of British Columbia. (2021, April 15). https://energystepcode.ca/compliance-tools-part9/.

PySimpleGUI. (n.d.). https://pysimplegui.readthedocs.io/en/latest/

PDFPlumber. (n.d.) GitHub. https://github.com/jsvine/pdfplumber

Camelot. (n.d.) https://camelot-py.readthedocs.io/en/master/

Pyinstaller. (n.d.) https://pyinstaller.readthedocs.io/en/stable/