Energy Performance Certificate Parser Tool

No. 9

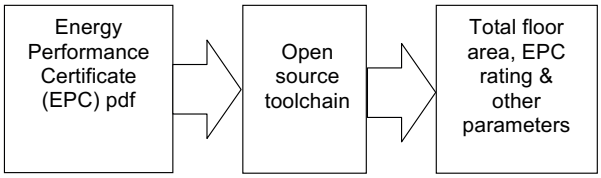
Luke Blunden, Phil Wu and AbuBakr Bahaj
(University of Southampton Liveable Cities Team)

Liveable Cities Tool

Bulk data on Energy Performance Certificates (EPC) were made freely available by the UK Government in March 2017. Previously EPCs in bulk were only available to purchase. Although this is a positive step, the bulk data often misses out, or incorrectly records key parameters which are nevertheless present in the original digital (Adobe PDF) version of the EPCs (for an example, see fig. 1). This parser reads an EPC, recording as many details as possible. It can handle scanned as well as original PDFs. The tool can be used to generate city level datasets for energy studies, or to validate or clean bulk EPC data.

Tool Contents

The user specifies an Energy Performance Certificate, runs the script and the results are output in a structured text file.



How has it been delivered?

The script is written in Bash and is available in both standalone form and as an online tool from November 2017 (hosted CGI script at energyandcities.org/epc2txt)

Figure 1: Example of a mis-recorded EPC parameter in UK database that is correctly identified by the parser script

Where has it been published?

The source code is available via Github at https://github.com/lsb1-soton/epc2txt and is distributed under the GNU General Public Licence v3.0. This code depends on other packages that may be licensed under different open source licenses.

Who participated?

The script was written by Luke Blunden with support from other members of the University of Southampton Liveable Cities Team. Supported by EPSRC grant ref EP/J017698/1.

Levels of Usability/Testability

The script depends on the poppler PDF rendering library (https://poppler.freedesktop.org/), Tesseract OCR (https://github.com/tesseract-ocr) and GhostScript (https://www.ghostscript.com/). It has been tested using Bash version 4.4.12 under Cygwin 2.8.2, using hundreds of EPCs from Southampton.

Liveable Cities · University of Birmingham · Lancaster University · UCL · University of Southampton · EPSRC