

officefileinfo

officefileinfo is a python script to help analyse the newer Microsoft Office file formats. There are numerous tools for dealing with the older OLE based files, but a distinct lack for the newer formats.

The newer formats all use the zip file format as the primary structure, with the internal sub formats using XML. As the primary structure is a zip file, it can obviously be trivially extracted, however, looking at each internal file is laborious and time consuming.

The aim of the script is to extract the pertinent information such as: does it contain macros, embedded objects etc, and also extract any macro code into separate files. The script uses the [officedissector](#) python library to do the file parsing.

Installation

The following steps show how to get step up using the Ubuntu distro:

```
wget https://bootstrap.pypa.io/get-pip.py
python get-pip.py
sudo apt-get install python
sudo pip install lxml
sudo pip install officedissector
```

Next either download the latest release archive or use git to clone the source code:

```
git clone https://github.com/infoassure/officefileinfo.git
```

Usage

To use the script, supply a Microsoft Office file such as a *.docx*, *.xlsx* and an output directory. The file will be parsed and it's various properties output to STDOUT. If any macro's are encountered, then they will be extracted and output to the directory supplied. Each macro will be in a separate file.

```
usage: officefileinfo.py [-h] -f FILE -o OUTPUTDIR
```