metadata harvester program

# User Manual

## Produced for University of Southern Queensland



Table of Contents

[User Manual 1](#_Toc22500553)

[Produced for University of Southern Queensland 1](#_Toc22500554)

[Introduction 3](#_Toc22500555)

[Features 3](#_Toc22500556)

[hardware Requirements 3](#_Toc22500557)

[Software Requirements 3](#_Toc22500558)

[Metadata Harvester Usage 4](#_Toc22500559)

[Command Line Arguments 4](#_Toc22500560)

## Introduction

Our Metadata Harvester script is a simple yet verbose python 3 script which can traverse directory trees from various operating systems for metadata on various file types. Bundled with the Harvester script is also the Metadata Searcher companion software, which utilises a cross-platform QT framework to visualise the metadata extracted from the harvester script.

The Harvester script is invoked from the command line via the python3 interpreter and passed multiple arguments based upon the user’s requirements.

### Features

* Ability to set depth of folder traversal by choosing between command line arguments for recursive folder traversal or non-recursive.
* Output file metadata into a simplified CSV (Comma Separated Value) file for use with GUI
* Useful help menu, easily accessible with the -h or –help command line flag
* Ability to parse metadata from the vast majority of popular file types

### hardware Requirements

* Processor: Intel Pentium 4 1.3GHz / AMD Athlon XP 1600+
* Graphics: N/A – Graphics provided by CPU will suffice
* Memory: 512mb RAM
* Storage: 500mb HDD or SSD space

### Software Requirements

* Operating System: Unix, Windows, MacOS
* Python 3.7.3 or higher
  + External Python Libraries Required:
    - Hachoir 3.0a5 or higher
    - Hachoir-core 1.3.3 or higher
    - Hachoir-metadata 1.3.3 or higher
    - Hachoir-parser 1.3.4 or higher
    - Python-magic 0.4.15 or higher
    - PyPDF2 1.26.0 or higher
    - PyQt5 5.13.1 or higher
    - Pip3 or higher

N.B – Python library requirements can be auto-installed by using our command line requirements installation tool.

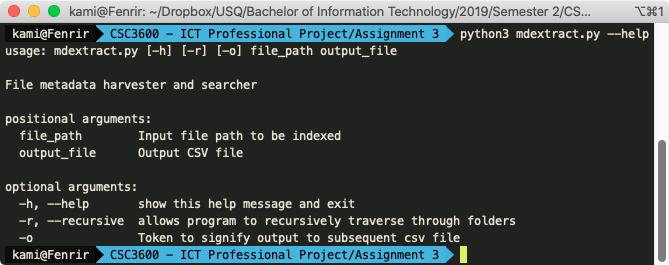
## Metadata Harvester Usage

The Harvester script is invoked via a terminal emulator (on unix/linux systems) or the command prompt (on Windows systems), through the Python 3 interpreter. An example execution is provided below without any command line arguments:



With no command line arguments entered the script will notify the user of an error and request any arguments that are considered requirements for execution of the script.

For a comprehensive list of all command line arguments the help menu can be invoked via the -h or –help options:



### Command Line Arguments

The positional (required) arguments and a brief description of each are as follows:

#### file\_path

This is the file path (in either relative of absolute form) to the input file that is to be indexed by the script.

An example file path can be:

**Simplified:**

/Users/kami/Desktop/test

**With escaped characters:**

/Users/kami/Dropbox/USQ/Bachelor\ of\ Information\ Technology/2019/Semester\ 2/CSC3600\ -\ ICT\ Professional\ Project/Assignment\ 3/test

#### Output\_File

This is the output csv file path (with extension) where formatted metadata will be stored.

The format is exactly the same as the file\_path argument with the exception of an added file extension .csv.

The optional arguments and a brief description of each are as follows:

#### help

The help menu can be accessed with:

* -h or - -help

The help menu has already been discussed above.

#### Recursive

The recursive option allows the Metadata Harvester script to fully traverse all subfolders from the root file\_path that was specified by the user. If the recursive option is not present then the script will scrape metadata in a non-recursive manner by default.

#### Output

The output argument facilitates the outputting of scraped metadata to a CSV file.

A fully constructed command would look as follows:



## Notes

The format of the command line arguments must be as follows:

Python3 mdextract.py [-r] file\_path [-o] output\_file

Any other misconfigurations may result in unwanted behaviour as the script has been designed to accommodate a particular format.