Package statistics

# Introduction

The main purpose of the module, hence the command line tool, is to retrieve from a specified address an archive file, which is actually a Contents index file associated with a \*.deb package.

The tool should be able to download the compressed Contents file, unzip it and then it should parse the file and output the statistics of the top 10 packages that have the most files associated with them.

# Structure

The tool has been thought out initially as a script, to support running it singularly, however instead of the standard functional approach we chose to use an object oriented one, since it would have more added value for scalability in and additional extensions in the future.

Therefore, a PackageContentsHandler class has been created, which in turn is the one that allows us to perform certain operations on the Contents index file. It does receive parameters for architecture and package address, which allows it to operate untied to specific hardcoded values if necessary, and can be used for further processing and extensions aside from the current implementation.

The package handler class has a specific function used for the retrieval of the Contents index compressed file and its unzipping process, as well as specific functions for the building process of the statistics and printing the in both a written report and console output.

# Usage

The module itself can be used as a standalone command line tool, as well as imported into a larger scale project where we could make use of the package contents handler class for further processing.

Example of usage as a command line tool:

