

Forum:[Hack Forums](#)

Thread: [Multiple Text Document Searching](#)

dwaynehicks - July 18, 2022, 10:21 am

Hello... I am researching a large breached data set that is made up of multiple tar.gz files that when extracted are plain text documents with the data. The total data set is in excess of 30GB.

If I am looking for specific data (say an email address) what is the best way to search these files? They are all in one folder, so is there a script or other program I could run for this task? Apologies if this is a super dumb question!! ?

Malthemar - July 18, 2022, 10:45 am

(Jul 18, 2022 10:21 AM)dwaynehicks Wrote: [\[CITE\]](#) Hello... I am researching a large breached data set that is made up of multiple tar.gz files that when extracted are plain text documents with the data. The total data set is in excess of 30GB.

If I am looking for specific data (say an email address) what is the best way to search these files? They are all in one folder, so is there a script or other program I could run for this task? Apologies if this is a super dumb question!! ?

You could have a script written for you that recursively opens each file in a directory, searches for text that matches a Regular Expression, saves the detected text elsewhere, and moves the processed file to a different directory.

By the end, you would have the original files untouched, and a list of emails (or whatever else you wanted to search for).