

# README.DOCX

## Dirty word Searcher



**Note:** Prior to using the tool, it's essential to thoroughly review and comprehend its requirements, functionalities, and limitations. Any misuse of the tool leading to unexpected outcomes will be the responsibility of the operator.

## Contents

<a href="#">1. Overview</a>	<a href="#">3</a>
<a href="#">2. Prerequisites</a>	<a href="#">3</a>
<a href="#">3. Configuration</a>	<a href="#">3</a>
<a href="#">4. File types</a>	<a href="#">3</a>
<a href="#">5. Running Dirty Word Searcher</a>	<a href="#">4</a>

## 1. Overview

The Dirty Word Searcher is a script designed to scan files on your computer for predefined "dirty" or inappropriate words. This can be useful for content moderation, filtering, or auditing text files for certain words.

The program offers various fields and options that users must select to maintain control over the targeted artefacts and keywords of interest.

## 2. Prerequisites

- Windows operating system

This executable is designed for Windows-based file systems and does not require any external dependencies beyond the standalone executable.

The program has not been tested on Linux/Unix-like systems, and the team does not guarantee its performance or the accuracy of results on those platforms.

## 3. Configuration

- a. Create a Dirty Words List

Create a text file to include the words and phrases you want to search for, with each entry on a new line. Ensure that the file has a .txt extension. The program will only match exact words and phrases, so include all necessary variations.

- b. Set File Paths for Search and Output Directories

Configure the script to scan the appropriate directories. You can manually enter these paths or use the browse buttons.

**Note:** The input requires at least a drive letter (e.g., "C:") and cannot handle multiple drives simultaneously. This limitation is due to performance considerations and the volume of data the program can process at one time. If your device has multiple drives, run the program separately for each drive.

## 4. File types

The Dirty Word Searcher supports the following file types:

- **Documents:** Excel, Word, PowerPoint, PDF, TXT

These file types will be identified, but the program will not search for dirty words within them, as this feature is not supported. They are categorised under **"Other types"**.

- **Images:** JPEG, PNG, BMP, TIFF
- **Videos:** MP3, MP4, AVI, MOV
- **Emails:** MSG

However, due to limited accuracy in scanning movies and images for dirty words, these files are highlighted for user review and are not processed through the dirty words text files.

**Note:** When searching for ".txt" files from the root directory (e.g., "C:"), the large volume of system documentation may increase the chances of false positives and potentially cause the program to crash, depending on the dirty word list provided.

## 5. Running Dirty Word Searcher

There are a number of user inputs required by the user. Users must have a valid txt file for the dirty words file, have a valid starting and output path.

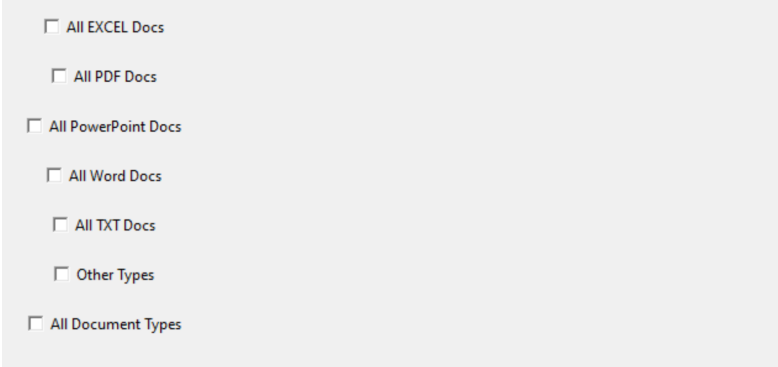
- a. Set File Paths. Ensure the script is configured to scan the appropriate directories and output the results.



The screenshot shows the 'Dirty Word Searcher' application window. It has a title bar with a feather icon and the text 'Dirty Word Searcher'. Inside the window, it says 'WELCOME TO DIRTY WORD SEARCHER!!!! :)'. Below this, there are three input fields: 'Dirty Word Text file:', 'Starting Directory for your search:', and 'Output Directory for your results:'. To the right of each input field is a 'Browse' button. At the bottom of the window are three buttons: 'Search', 'Reset', and 'Quit'.

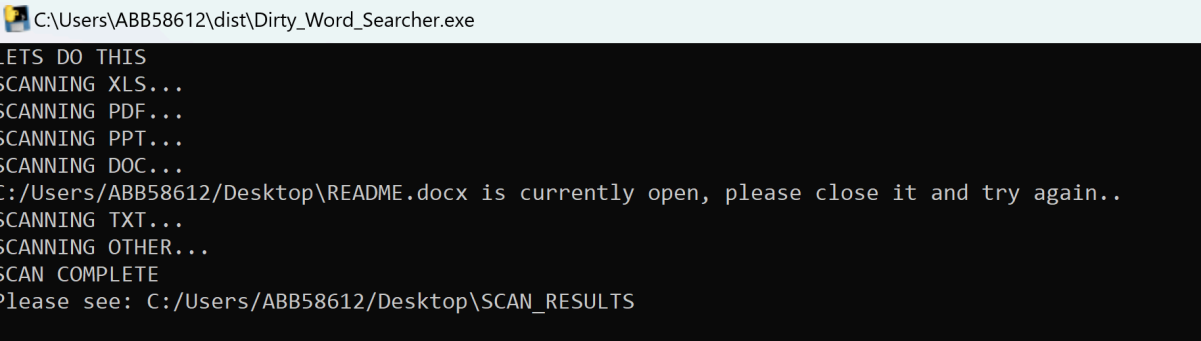
- b. The checkboxes down the left hand side give the users power to determine which artefacts they wish to interrogate.

The **“Other Types”** option will search for all other file types to pass through the dirty word searcher.



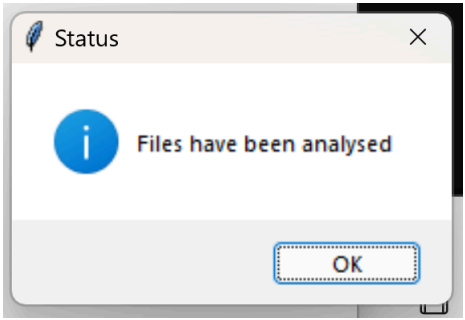
The screenshot shows a list of checkboxes for file types. The options are: 'All EXCEL Docs', 'All PDF Docs', 'All PowerPoint Docs', 'All Word Docs', 'All TXT Docs', 'Other Types', and 'All Document Types'. All checkboxes are currently unchecked.

- c. The Command Prompt window in the background will display messages and updates on the scan's progress.



```
C:\Users\ABB58612\dist\Dirty_Word_Searcher.exe
LETS DO THIS
$CANNING XLS...
$CANNING PDF...
$CANNING PPT...
$CANNING DOC...
C:/Users/ABB58612/Desktop\README.docx is currently open, please close it and try again..
$CANNING TXT...
$CANNING OTHER...
$CAN COMPLETE
Please see: C:/Users/ABB58612/Desktop\SCAN_RESULTS
```

- d. After the scan completes, a message box will appear saying “Files have been analysed”.



- e. Review the contents of the results in the designated output directory. Detected instances of dirty words will need manual confirmation to ensure accuracy, as this program has not undergone extensive testing to guarantee high confidence levels.

Name	Date modified	Type	Size
Other_Files	28/05/2024 2:52 PM	File folder	
Word_Positive_Results	28/05/2024 2:52 PM	File folder	
Excel_Cannot_Anayse_List.out	28/05/2024 2:52 PM	OUT File	1 KB
Excel_Positive_Results.out	28/05/2024 2:52 PM	OUT File	1 KB
PDF_Cannot_Anayse_List.out	28/05/2024 2:26 PM	OUT File	1 KB
PDF_files_without_text.out	28/05/2024 2:52 PM	OUT File	1 KB
PDF_Positive_Results.out	28/05/2024 2:52 PM	OUT File	2 KB
Powerpoint_Cannot_Anayse_List.out	28/05/2024 2:52 PM	OUT File	1 KB
Powerpoint_Positive_Results.out	28/05/2024 2:52 PM	OUT File	1 KB
TXT_Cannot_Anayse_List.out	28/05/2024 2:52 PM	OUT File	1 KB
TXT_Positive_Results.out	28/05/2024 2:52 PM	OUT File	1 KB