Google Test Data Sanitization

**TABLE OF CONTENTS**

[PREREQUISITES](#h.6qhhqra0kfvr)

[SOURCE CODE LOCATION](#h.qlyearuookp)

[CODE ORGANIZATION/ARCHITECTURE](#h.piuivkvboshl)

[HOW TO SETUP](#h.hnrfkimhoy9)

[TEST DATA USED](#h.bv3dbt4s5yw6)

[LINK TO TOOL IN PRODUCTION](#h.7nuq2qhksqbc)

[NOTES](#h.mxw7gcev1mhl)

# **PREREQUISITES**

It runs on Windows .But for required resolution run it on :

* Platform: Windows 7 – 64 bit.
* Microsoft office 2007
* Dot net framework 4.5

**SOURCE CODE LOCATION**

scratch/dotNetTools : [PUNETETools/dotNetTools/Google\_TE\_O2O/TestDateSanitization/](https://quickoffice-internal.googlesource.com/html-office/+/scratch/dotNet/PUNETETools/dotNetTools/Google_TE_O2O/TestDateSanitization/)

# **CODE ORGANIZATION/ARCHITECTURE**

# **HOW TO SETUP**

* Install Dot net framework 4.5 and office 2007.
* Clone the code from scratch TETools branch.
* Get Corruption handler EXE from -

“PUNETETools\dotNetTools\Google\_TE\_O2O\OfficeFilesValidityChecker

\bin\Release\ OfficeFilesValidityChecker.exe ”

* Get Test Sanitization File From –

“PUNETETools\dotNetTools\Google\_TE\_O2O\TestDateSanitization\

Test\_data\bin\Release\Test\_data.exe”

* Get Watermarked for Docx EXE from –

“PUNETETools\dotNetTools\Google\_TE\_O2O\TestDateSanitization\Watermarked \_sanitization\

Watermarked for Docx\bin\Release\Watermarked for Docx.exe”

# **HOW TO RUN**

* Run Corruption handler on all files:-

- OfficeFilesValidityChecker.exe <-input folder-> <-output folder->

* Run test sanitization tool on valid files from corruption handler Output.

- Test\_data.exe. <-input folder-> <-output folder->

* For watermarked count from docx run on acceptable files form test sanitization tool.

- Watermarked for Docx.exe. <-input folder-> <-output folder->

# **TEST DATA USED**

# **LINK TO REPORT**

* [Test Data Sanitization](https://docs.google.com/a/synerzip.com/spreadsheet/ccc?key=0ArCsyuFa4cI8dEM5X2RYeURKUjBxcXN3RXB4TnZiNUE&usp=drive_web#gid=4)

# **NOTES**

* **Limitations :**

1. Unable to get water marked form 2k3 files.
2. Unable to get all text from sheet in small time.