

Welcome to the NIST/CFTT String Search Data Set. This is Version 1.0 of test data for string searching with Federated Testing.

This data set is intended to be used with Federated Testing Version 4.0 and later for string search testing.

Copy the folder named, “copy-to-test-computer” to the computer where the search tool you want to test is installed. The other files are for reference if you want to look deeper at the image files, but these files (xxx_string_doc.txt and xxx_hit-dump.txt) can be located anywhere.

File Inventory:

FILE	Content
read-me-first.rtf	This File
ss-win-03-06-18.dd	Image file for Windows. This file should be attached, imported or added to the search tool as evidence. This is the Windows data set.
win_string_doc.txt	Location of target strings in Windows dd file. Listed by string ID.
win_hit-dump.txt	Hex dump around each target string in Windows dd file.
ss-unix-03-06-18.dd	Image file for UNIX. This file should be attached, imported or added to the search tool as evidence. This is the UNIX data set.

unix_string_doc.txt	Location of target strings in UNIX dd file. Listed by string ID.
unix_hit-dump.txt	Hex dump around each string in UNIX dd file.
string-search-test-cases.html	Specification of each test case, including search tool settings, the search string and a brief test case description. For non-English search strings, the string can be copied from the html file and pasted into the search tool.

The dd files were created as follows:

- 1) A set of base strings were selected.
- 2) For each base string, a set of files were created with the following characteristics:
 - i) Create two files for each partition and one file for unallocated space.
 - ii) The file name has the form:
STATUS-AltName-Partition-Encoding.txt
 - (1) STATUS is either LIVE or DELETED reflecting the final status of the file.
 - (2) AltName is a word or two words with similar meaning to the base string, e.g., the string WOLF has AltName of AllCap-Lupus. Two files used in the meta-data test case use the base string instead of an AltName (this gets the base string into meta-data, but with no string ID).
 - (3) Partition is where the file is located. One of these: fat, exfat, ntfs, unalloc, osxj, osxc or ext4.
 - (4) Encoding is the character encoding of the base string. One of

these: ascii, utf-8, utf-16be or utf-16le.

iii) File content begins with: "TESTFILE:(**FILE NAME**)"

iv) A list of random filler words with a nautical theme, e.g., creek, sea, river, tuna, squid, etc.

v) The base string (with a unique string ID for each file), formatted as follows:

encoding ==> STRING ID <=== partition

vi) More random nautical filler words.

vii) The file ends with:

Encoding (**FILE NAME**) END-OF-FILE

3) Files are copied to destination partitions on a zero-wiped 2GB drive.

4) Any special case files are copied to the drive, e.g., "lost files."

5) Files with STATUS of DELETED are deleted.

6) The drive is imaged.

7) The image file (.dd) is scanned for each string used as a search target in a test case. The files xxx_string_doc.txt and xxx_hit-dump.txt are created to document the location of each search target.

That's all, email CFTT@NIST.GOV if you have questions or need help.