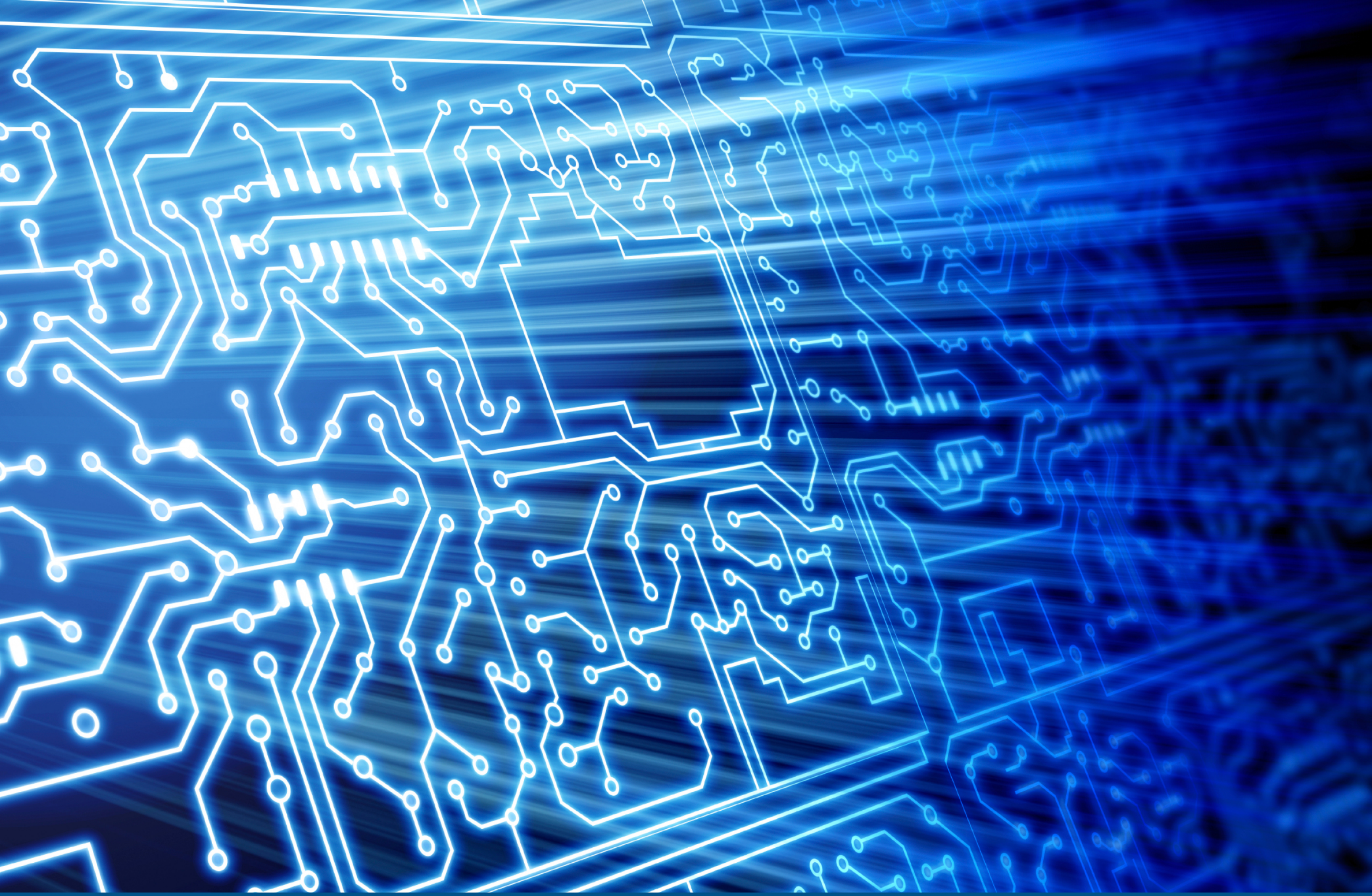


X-Ways Forensics Version 19.6-SR-4 x64

Test Results for String Search Tool

March 27, 2019

Homeland Security

Science and Technology

This report was prepared for the Department of Homeland Security Science and Technology Directorate Cyber Security Division by the Office of Law Enforcement Standards of the National Institute of Standards and Technology.

For additional information about the Cyber Security Division and ongoing projects, please visit http://www.dhs.gov/science-and-technology/cyber-security-division.

March 2019

Test Results for String Search Tool:
X-Ways Forensics Version 19.6-SR-4 x64

Introduction

The Computer Forensics Tool Testing (CFTT) program is a joint project of the Department of Homeland Security (DHS), the National Institute of Justice (NIJ), and the National Institute of Standards and Technology (NIST) Special Programs Office and Information Technology Laboratory (ITL). CFTT is supported by other organizations, including the Federal Bureau of Investigation, the U.S. Department of Defense Cyber Crime Center, U.S. Internal Revenue Service Criminal Investigation Division Electronic Crimes Program, and the U.S. Department of Homeland Security's Bureau of Immigration and Customs Enforcement, U.S. Customs and Border Protection and U.S. Secret Service. The objective of the CFTT program is to provide measurable assurance to practitioners, researchers, and other applicable users that the tools used in computer forensics investigations provide accurate results. Accomplishing this requires the development of specifications and test methods for computer forensics tools and subsequent testing of specific tools against those specifications.

Test results provide the information necessary for developers to improve tools, users to make informed choices, and the legal community and others to understand the tools' capabilities. The CFTT approach to testing computer forensics tools is based on well-recognized methodologies for conformance and quality testing. The CFTT approach tests features that forensic labs are likely to use on a regular basis. Interested parties in the computer forensics community can review and comment on the specifications and test methods posted on the CFTT Web site (https://www.cftt.nist.gov).

This document reports the results from testing the string search function of X-Ways Forensics Version 19.6-SR-4 (http://www.x-ways.net/) using the CFTT Federated Testing Test Suite Version 4.0 (beta version, final to be released in 2018) using String Searching data set Version 1.1.

Federated Testing is an expansion of the CFTT program to provide forensic investigators and labs with test materials for tool testing and to support shared test reports. The goal of Federated Testing is to help forensic investigators to test the tools that they use in their labs and to enable sharing of tool test results. CFTT's Federated Testing Forensic Tool Testing Environment and included test suites can be downloaded by visiting https://www.cftt.nist.gov and selecting Federated Testing. The results can be optionally shared with CFTT, reviewed by CFTT staff, and then shared with the community.

Test results from this and other tools can be found on DHS's computer forensics web page, https://www.dhs.gov/science-and-technology/nist-cftt-reports.

Table of Contents

How to Read This Report

This report is organized into these sections:

1. Tested Tool Description: The tool name, version, and vendor information are listed.
2. Results Summary: This section identifies any significant anomalies observed in the test runs. This section provides a narrative of key findings identifying where the tool meets expectations and provides a summary of tool behaviors that did not meet expectations.
3. Test Environment & Selected Test Cases: Description of hardware and software used in tool testing and a list identifying the applicable test cases from the Federated Testing String Search Test Suite.
4. Test Result Details by Case: Automatically generated test results that identify anomalies.
5. Search Configuration Settings: Screen captures of sample search configurations selected during testing.

Test Results for String Search Tool: X-Ways Forensics Version 19.6-SR-4 x64

1 Tested Tool Description

Tool Name: X-Ways Forensics
Tool Version: Version 19.6-SR-4 x64
Vendor: X-Ways Software Technology AG

2 Results Summary

The test data set and test cases used to create this test report are limited to frequently encountered aspects of searching for text. Trying to cover every feature is not practical, but these test cases do cover a broad range of features. The features that are addressed in the full test data set (including features that X-Ways Forensics does not support) are listed below:

- File System: MS Windows (FAT, exFAT, NTFS) and UNIX-like (Ext4, OSXJ -- Mac OS Extended (Journaled), OSXC -- Mac OS Extended (Case-sensitive, Journaled) and APFS – Apple File System).
- String Location: Active File, Deleted (but recoverable) file, Unallocated Space, and Meta-Data.
- Search Method (aka search engine): Indexed, Live or Physical.
- String Encoding: ASCII, UTF-8, UTF-16BE and UTF-16LE with and without a **byte order mark**.
- Normalized Unicode: Match alternative forms of character representation, e.g., the substring “if” of the string “infinity” could be represented by a single ligature character or two separate characters, a letter with a diacritic mark could be represented by either one or two characters. A search for any one representation should match either representation. See section 4.3.
- Language: In addition to English, strings representative of diacritical marks (German, French, Spanish), non-Latin characters (Russian), right-to-left presentation (Arabic), and Asian languages (Chinese, Japanese and Korean).
- Fragmented File: String that spans two disjoint file fragments.
- Logical Operations: Combine search results with logical operators **and**, **or** and **not**.
- Stemming: Match inflected forms derived from a word stem, e.g., a search for *run* should also match *runs*, *running* and *ran*.
- Embedded Formatting: String with embedded formatting. MS Word and HTML.

The following features are not supported by X-Ways Forensics:

- Normalized Unicode string searching.
- Stemming search.
- Apple File System is not supported, but it is treated as unallocated space.

Three search engines were tested: Live Search, Physical Search and Indexed Search. In general, live and physical search exhibited few unexpected behaviors, however indexed search exhibited several anomalies.

2.1 General Observations

Determining if Unicode UTF-16 text is UTF-16-BE or UTF-16-LE is problematic for some text samples, especially for Latin based characters, because a one-byte shift in starting point for a string can align with either representation. For example, consider the hex representing the string “Sch nheit” in UTF-16:

```
00 53 00 63 00 68 00 f6 00 6e 00 68 00 65 00 69 00 74 00
     S     c     h     o:    n     h     e     i     t
```

If you start the match with 00 53 00 63 00 . . . then it is UTF-16-BE, but
If you start the match with 53 00 63 00 . . . then it is UTF-16-LE, so without any other information is could be either BE or LE. This is an artifact of UTF-16 characters that have a first byte of zero for the big-endian representation (as in Latin based characters).
The tool designer has several choices for reporting a match of UTF-16, for X-Ways, the designer chose to report any matches encountered. This will appear to the tool user as the same UTF-16 string apparently being reported twice for some strings (those with all zero bits in the first byte of a big-endian representation) and not for other strings (those without a zero first byte such as Cyrillic, Arabic or Asian).
This is not a problem with the tool, but the tool user should be aware of the behavior.

2.2 Live Search Anomalies and Observations

The following behaviors were observed when using the live (logical) search engine:

- No normalization of search strings is performed. Some Unicode strings may be in one of several possible normalized forms. Each form must be explicitly searched for. See section 4.3.
- Unicode UTF-16 strings of Latin text are reported twice. However, other non-Latin text (Arabic, Russian, Chinese, etc.) is only reported once.
- HTML text with embedded formatting tags in an APFS file system or in unallocated space is not reported.
- Target strings in MS Word DOCX files located in unallocated space are not reported. We may have not configured the tool to correctly search DOCX files in unallocated space.

2.3 Physical Search Anomalies and Observations

The following behaviors were observed when using the physical search engine:

- FT-SS-06 searches for instances of the string “fox” without a nearby occurrence of the string “tiger” yielded a “page fault.”
- No normalization of search strings is performed. Some Unicode strings may be in one of several possible normalized forms. Each form must be explicitly searched for. See section 4.3.
- Unicode strings in the test image that are normalized as NFC are reported twice.
- Unicode UTF-16 strings of Latin text are reported twice. However, other non-Latin text (Arabic, Russian, Chinese, etc.) is only reported once.

The following behaviors are built-in limitations of a physical search.

- Container formats such as MS Windows DOCX are not decoded and so the strings “nitroglycerin” and “peroxide” are not reported for test case FT-SS-09-DOC.
- Because logical file structure is not examined in a physical search, the string “Washington” which is split across two noncontiguous file fragments in test case FT-SS-09-Frag is not reported.

2.4 Indexed Search Anomalies and Observations

Indexed search is a two-step process. First, an index must be built before any searches are run. Second, search targets are looked-up in the index. The following behaviors were observed when using the indexed search engine:

- Building an index presents a choice of either the “Unicode Multi-Lingual Plane” or a specific language. Selecting the “Multi-Lingual Plane” successfully builds an index but no matches are reported when searching.
- Selecting “Japanese” language generates an error message and the index creation fails.
- No normalization of search strings is performed. Some Unicode strings may be in one of several possible normalized forms. Each form must be explicitly searched for. See section 4.3.
- Searching Korean (Hangul) text crashed the tool.
- HTML text with embedded formatting tags in an APFS file system or in unallocated space is not reported.
- Target strings in MS Word DOCX files located in unallocated space are not reported. We may have not configured the tool to correctly search DOCX files in unallocated space.
- Unicode UTF-16 strings of Latin text are reported twice. However, other non-Latin text (Arabic, Russian, Chinese, etc.) is only reported once.
- Unicode UTF-16-BE Arabic or Russian strings are not reported.

- Existing meta-data copies of the string *caó n* was not reported for FAT32 and ExFAT file systems.

3 Test Environment & Selected Test Cases

This section describes test hardware, software, test data sets and test cases.

3.1 Test Hardware and Software

X-Ways Forensics Version 19.6-SR-4 x64 was installed on a Dell OptiPlex 7050 with 32GB installed RAM, running Microsoft Windows 10 Enterprise, Version 1607, OS Build 14393.2068.

Testing was performed using CFTT Federated Testing Test Suite Version 4.0 (beta version, final to be released in 2018).

3.2 Test Data Sets and Test Cases

This section discusses the test data sets and the test cases used in testing.

3.2.1 Test Data Sets

String search test data set package Version 1.1 was used. The package can be downloaded from either the CFTT web site (go to www.cftt.nist.gov and then select String Searching) or the CFReDS web site (www.cfreds.nist.gov). The package includes two dd files with known content. One of the dd test images contains target strings within FAT, ExFAT and NTFS file systems (Windows), the other dd test image contains target strings from HFS+ journaled, case insensitive (OSXJ), HFS+ journaled, case sensitive (OSXC), ext4 file system and APFS (Apple file system) (UNIX-like).

In general, each target string is encoded in ASCII and located in both an active file and a recoverable deleted file in each partition of the test image. The Windows dd image also has a block of unallocated storage that contains the target strings without a file system. Some of the target strings are also encoded in Unicode UTF-8, UTF-16BE and UTF-16LE with a byte-order-mark. Test case FT-SS-07 is organized to test language and Unicode specific situations such as Unicode UTF-16 without a byte-order-mark, Unicode text with and without combining characters (diacritic marks), Unicode text with and without ligatures ("fi" as two characters and as one character) Test case FT-SS-09 is organized to test specific situations such as formatted strings, strings spanning file fragments, and strings located in inaccessible areas. Each instance of a target string also has a unique associated string ID located immediately after the target string. The string ID helps identify the specific string matched by the search tool.

3.2.2 Test Case Descriptions

The following table gives a brief description of available test cases in the data sets. Not all test cases are used for all data sets.

Table 1 Test Cases

Case	Case Description
FT-SS-01	Search ASCII
FT-SS-02	Search Ignore Case
FT-SS-03	Search for Words
FT-SS-04	Search Logical AND
FT-SS-05	Search Logical OR
FT-SS-06	Search Logical NOT
FT-SS-07-CJK-char	Search Unicode Chinese/Japanese ideograms (Asian)
FT-SS-07-CJK-Hangul	Search Unicode CJK Korean Hangul (Asian)
FT-SS-07-CJK-kana	Search Unicode CJK Japanese phonetic Kana (Asian)
FT-SS-07-Cyrillic	Search Unicode Cyrillic (Russian)
FT-SS-07-Latin	Search Unicode Latin (French & German)
FT-SS-07-NoBOM	Search Unicode 16 without a byte-order-mark
FT-SS-07-Norm	Search Unicode 16 for normalized diacritic marks (NFC & NFD) and ligatures (NFKC & NFKD)
FT-SS-07-RTL	Search Unicode RTL (Arabic)
FT-SS-08-Email	Search Tool-defined Queries -- Email Address
FT-SS-08-Phone	Search Tool-defined Queries -- Telephone Number
FT-SS-08-SS	Search Tool-defined Queries -- Social Security
FT-SS-09-Doc	Search Formatted Document Text
FT-SS-09-Frag*	Search Fragmented File
FT-SS-09-Lost*	Search Inaccessible (lost) Areas
FT-SS-09-MFT*	Search File in NTFS MFT
FT-SS-09-Meta	Search file name substring in Meta-data
FT-SS-09-Stem	Search for matches to word stem
FT-SS-10-Hex	Search Hexadecimal Character Match
FT-SS-10-Regex	Search Pattern Character Match

Some test cases are for specific features, e.g., logical conditions (**and**, **or**, **not**), built in searches (email, telephone numbers), etc. Three test cases (marked with "*"), FT-SS-09-Frag, FT-SS-09-Lost & FT-SS-09-MFT are only applied to the Windows data set.

4 Test Result Details by Case (per Data Set)

A string search tool may implement more than one search algorithm (also known as a search engine) for searching text. The two most common search engines are *indexed search* and *live search*. An indexed search reads all the acquired data once before doing any searching and builds an index to all words found. Each query can be looked up quickly in the index. A Live search reads all the acquired data for each query. In addition, some tools such as X-Ways also support a sequential scan of the acquisition by sectors. This is called a *physical search* and has some limitations such as missing strings split across cluster fragment boundaries and missing and strings that must be extracted from a file such as an archive, etc.

This section presents test results by test image: Windows file systems, or UNIX-like file systems. For each test image, there is a result table for each search engine tested. Each table shows results by test case of the number of expected search hits, the number of actual search hits and the number of strings missed (i.e., expected hits minus actual hits) for allocated files, deleted files and unallocated space.

The following search engines were tested: Physical, Indexed and Live.

4.1 Results for Data Set: Windows

This section provides results for the Windows data set.

4.1.1 Results for Physical Search of Windows Data Set

The table columns contain the following information:

- **Case:** The test case identifier.
- **Expected String:** The strings that should be reported by the search.
- **Active Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in an active file.
- **Deleted Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in a deleted file.
- **Unallocated Space:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in unallocated space.
- **Expected:** The number of instances of the expected string found in the group (i.e., Active files, Deleted files or Unallocated space).
- **Hits:** The number of times the expected string was found in the group.
- **Misses:** The number of times the expected string was missed (not found) in the group.

Notes: If the row identifies a test case, then the results are a summary for all the strings that should be found.

In the Expected String column for test case FT-SS-09-DOC each string is labeled to indicate features of the expected string. The labels include the file type (.doc, .docx or .html) and the encoding of the string (if a .doc file). If the string has embedded formatting it is labeled as *Formatted*, e.g., the string *crossbow* has the substring *cross* formatted as bold and underlined, i.e., **cross**bow.

Results for Physical Search of Windows Data Set

Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-01	Summary	3	3	0	3	3	0	1	1	0
	DireWolf	3	3	0	3	3	0	1	1	0
FT-SS-02	Summary	15	15	0	15	15	0	5	5	0
	WOLF	3	3	0	3	3	0	1	1	0
	wolf	3	3	0	3	3	0	1	1	0
	Wolf	3	3	0	3	3	0	1	1	0
	DireWolf	3	3	0	3	3	0	1	1	0
	WereWolf	3	3	0	3	3	0	1	1	0
FT-SS-03	Summary	9	9	0	9	9	0	3	3	0
	WOLF	3	3	0	3	3	0	1	1	0
	wolf	3	3	0	3	3	0	1	1	0
	Wolf	3	3	0	3	3	0	1	1	0
FT-SS-04	Summary	3	3	0	3	3	0	0	0	0
	panda and fox	3	3	0	3	3	0	0	0	0
FT-SS-05	Summary	6	6	0	6	6	0	2	2	0
	DireWolf	3	3	0	3	3	0	1	1	0
	WereWolf	3	3	0	3	3	0	1	1	0
FT-SS-06	Summary	12	0	12	12	0	12	0	0	0
	fox and not tiger	12	0	12	12	0	12	0	0	0
FT-SS-07-CJK-char	Summary	18	18	0	18	18	0	6	6	0
	□ □	9	9	0	9	9	0	3	3	0
	□ □	9	9	0	9	9	0	3	3	0
	Summary	9	9	0	9	9	0	3	3	0

Results for Physical Search of Windows Data Set

Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-07-CJK-hangul	서울	9	9	0	9	9	0	3	3	0
FT-SS-07-CJK-kana	Summary	18	18	0	18	18	0	6	6	0
	スバル	9	9	0	9	9	0	3	3	0
	みつびし	9	9	0	9	9	0	3	3	0
FT-SS-07-Cyrillic	Summary	9	9	0	9	9	0	3	3	0
	Сибирь	9	9	0	9	9	0	3	3	0
FT-SS-07-Latin	Summary	18	18	0	18	18	0	6	6	0
	garçon	9	9	0	9	9	0	3	3	0
	Schönheit	9	9	0	9	9	0	3	3	0
FT-SS-07-NoBOM	Summary	39	39	0	39	39	0	13	13	0
	Россия	9	9	0	9	9	0	3	3	0
	فلافل	9	9	0	9	9	0	3	3	0
	□ □	9	9	0	9	9	0	3	3	0
	QuarterHorse	12	12	0	12	12	0	4	4	0
FT-SS-07-Norm	Summary	75	75	0	75	75	0	25	25	0
	mañana (NFD)	9	9	0	9	9	0	3	3	0
	infinity (No Ligature)	12	12	0	12	12	0	4	4	0
	Mäuse (NFD)	9	9	0	9	9	0	3	3	0
	infinity (Ligature)	9	9	0	9	9	0	3	3	0
	Mäuse (NFC)	9	9	0	9	9	0	3	3	0
	libertà (NFC)	9	9	0	9	9	0	3	3	0
	libertà (NFD)	9	9	0	9	9	0	3	3	0
	ma ana (NFC)	9	9	0	9	9	0	3	3	0
FT-SS-07-RTL	Summary	9	9	0	9	9	0	3	3	0
	الكسكس	9	9	0	9	9	0	3	3	0

Results for Physical Search of Windows Data Set

Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-09-Doc	Summary	16	13	3	0	0	0	16	13	3
	longbow .html	2	2	0	0	0	0	2	2	0
	shotgun Formatted .doc UTF-16	2	2	0	0	0	0	2	2	0
	revolver .doc UTF-16	2	2	0	0	0	0	2	2	0
	peroxide .docx	2	1	1	0	0	0	2	1	1
	nitroglycerin Formatted .docx	2	1	1	0	0	0	2	1	1
	rifle .doc UTF-8	2	2	0	0	0	0	2	2	0
	crossbow Formatted .html	2	1	1	0	0	0	2	1	1
	flintlock Formatted .doc UTF-8	2	2	0	0	0	0	2	2	0
FT-SS-09-Frag	Summary	2	1	1	0	0	0	0	0	0
	Washington	1	0	1	0	0	0	0	0	0
	California	1	1	0	0	0	0	0	0	0
FT-SS-09-Lost	Summary	0	0	0	0	0	0	4	4	0
	SecretKey	0	0	0	0	0	0	2	2	0
	disconnected	0	0	0	0	0	0	2	2	0
FT-SS-09-MFT	Summary	4	4	0	4	4	0	0	0	0
	bear	4	4	0	4	4	0	0	0	0
FT-SS-09-Meta	Summary	6	6	0	6	6	0	2	2	0
	cañón	3	3	0	3	3	0	1	1	0
	thunderbird	3	3	0	3	3	0	1	1	0
FT-SS-10-Hex	Summary	3	3	0	3	3	0	1	1	0
	panda	3	3	0	3	3	0	1	1	0

Results for Physical Search of Windows Data Set										
Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-10-Regex	Summary	6	6	0	6	6	0	2	2	0
	DireWolf	3	3	0	3	3	0	1	1	0
	WereWolf	3	3	0	3	3	0	1	1	0

4.1.2 Meta-Data results for Physical Search of Windows Data Set

The following table presents search results for strings located in file system meta-data. The **Case** column identifies the test case, the **String** column identifies the search string, the **Partition** column identifies the partition (file system) where the string is located and the **Seen** column records if the search tool reported at least one instance of the string (yes or no) in meta-data.

Meta-Data Results for Physical Search of Windows Data Set			
Case	String	Partition	Seen
FT-SS-09-Meta			
	thunderbird	ntfs	Yes
	cañón	fat32	Yes
	cañón	exfat	Yes
	cañón	ntfs	Yes

4.1.3 Comments on Physical Search of Windows Data Set

The following table presents any comments recorded during testing for a test case.

Case	Comments on Physical Search of Windows Data Set
FT-SS-06	Page Fault
FT-SS-07-Latin	UTF-16 encoded strings are reported twice.
FT-SS-07-NoBOM	Hits on the string "QuarterHorse" encoded as UTF-16 are reported twice.
FT-SS-07-Norm	Searches do not use Unicode normalization on the search string. Strings normalized as NFC are reported twice.
FT-SS-09-Frag	Not finding the string "Washington" is the real expected result because the string is split across two file fragments and should be missed in a sector by sector physical search.

4.1.4 Results for Indexed Search of Windows Data Set

The table columns contain the following information:

- **Case:** The test case identifier.
- **Expected String:** The strings that should be reported by the search.
- **Active Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in an active file.
- **Deleted Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in a deleted file.
- **Unallocated Space:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in unallocated space.
- **Expected:** The number of instances of the expected string found in the group (i.e., Active files, Deleted files or Unallocated space).
- **Hits:** The number of times the expected string was found in the group.
- **Misses:** The number of times the expected string was missed (not found) in the group.

Notes: If the row identifies a test case, then the results are a summary for all the strings that should be found.

In the Expected String column for test case FT-SS-09-DOC each string is labeled to indicate features of the expected string. The labels include the file type (.doc, .docx or .html) and the encoding of the string (if a .doc file). If the string has embedded formatting it is labeled as *Formatted*, e.g., the string *crossbow* has the substring *cross* formatted as bold and underlined, i.e., **cross**bow.

Results for Indexed Search of Windows Data Set										
Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-01		3	3	0	3	3	0	1	1	0
	DireWolf	3	3	0	3	3	0	1	1	0
FT-SS-02		15	15	0	15	15	0	5	5	0
	WOLF	3	3	0	3	3	0	1	1	0
	wolf	3	3	0	3	3	0	1	1	0
	Wolf	3	3	0	3	3	0	1	1	0
	DireWolf	3	3	0	3	3	0	1	1	0
	WereWolf	3	3	0	3	3	0	1	1	0
FT-SS-03		9	9	0	9	9	0	3	3	0
	WOLF	3	3	0	3	3	0	1	1	0

Results for Indexed Search of Windows Data Set										
Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
	wolf	3	3	0	3	3	0	1	1	0
	Wolf	3	3	0	3	3	0	1	1	0
FT-SS-04		3	3	0	3	3	0	0	0	0
	panda and fox	3	3	0	3	3	0	0	0	0
FT-SS-05		6	6	0	6	6	0	2	2	0
	DireWolf	3	3	0	3	3	0	1	1	0
	WereWolf	3	3	0	3	3	0	1	1	0
FT-SS-06		12	12	0	12	12	0	0	0	0
	fox and not tiger	12	12	0	12	12	0	0	0	0
FT-SS-07-CJK-char		18	18	0	18	18	0	6	6	0
	□ □	9	9	0	9	9	0	3	3	0
	□ □	9	9	0	9	9	0	3	3	0
FT-SS-07-CJK-hangul		9	9	0	9	9	0	3	3	0
	서울	9	9	0	9	9	0	3	3	0
FT-SS-07-CJK-kana		18	9	9	18	9	9	6	3	3
	スバル	9	0	9	9	0	9	3	0	3
	みつびし	9	9	0	9	9	0	3	3	0
FT-SS-07-Cyrillic		9	6	3	9	6	3	3	2	1
	Сибирь	9	6	3	9	6	3	3	2	1
FT-SS-07-Latin		18	15	3	18	15	3	6	5	1
	garçon	9	6	3	9	6	3	3	2	1

Results for Indexed Search of Windows Data Set										
Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
	Schönheit	9	9	0	9	9	0	3	3	0
FT-SS-07-NoBOM		39	27	12	39	27	12	13	9	4
	Россия	9	6	3	9	6	3	3	2	1
	فلافل	9	6	3	9	6	3	3	2	1
	□ □	9	3	6	9	3	6	3	1	2
	QuarterHorse	12	12	0	12	12	0	4	4	0
FT-SS-07-Norm		75	39	36	75	39	36	25	13	12
	mañana (NFD)	9	0	9	9	0	9	3	0	3
	libertà (NFD)	9	0	9	9	0	9	3	0	3
	Mäuse (NFD)	9	0	9	9	0	9	3	0	3
	infinity (Ligature)	9	0	9	9	0	9	3	0	3
	Mäuse (NFC)	9	9	0	9	9	0	3	3	0
	infinity (No Ligature)	12	12	0	12	12	0	4	4	0
	ma ana (NFC)	9	9	0	9	9	0	3	3	0
	libertà (NFC)	9	9	0	9	9	0	3	3	0
FT-SS-07-RTL		9	6	3	9	6	3	3	2	1
	الكسكس	9	6	3	9	6	3	3	2	1
FT-SS-09-Doc		16	15	1	0	0	0	16	13	3
	longbow .html	2	2	0	0	0	0	2	2	0
	shotgun Formatted .doc UTF-16	2	2	0	0	0	0	2	2	0

Results for Indexed Search of Windows Data Set										
Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
	revolver .doc UTF-16	2	2	0	0	0	0	2	2	0
	peroxide .docx	2	2	0	0	0	0	2	1	1
	nitroglycerin Formatted .docx	2	2	0	0	0	0	2	1	1
	rifle .doc UTF-8	2	2	0	0	0	0	2	2	0
	crossbow Formatted .html	2	1	1	0	0	0	2	1	1
	flintlock Formatted .doc UTF-8	2	2	0	0	0	0	2	2	0
FT-SS-09-Frag		2	2	0	0	0	0	0	0	0
	Washington	1	1	0	0	0	0	0	0	0
	California	1	1	0	0	0	0	0	0	0
FT-SS-09-Lost		0	0	0	0	0	0	4	4	0
	SecretKey	0	0	0	0	0	0	2	2	0
	disconnected	0	0	0	0	0	0	2	2	0
FT-SS-09-MFT		4	4	0	4	4	0	0	0	0
	bear	4	4	0	4	4	0	0	0	0
FT-SS-09-Meta		6	5	1	6	5	1	2	2	0
	cañón	3	2	1	3	2	1	1	1	0
	thunderbird	3	3	0	3	3	0	1	1	0
FT-SS-10-Hex		3	3	0	3	3	0	1	1	0
	panda	3	3	0	3	3	0	1	1	0

4.1.5 Meta-Data results for Indexed Search of Windows Data Set

The following table presents search results for strings located in file system meta-data. The **Case** column identifies the test case, the **String** column identifies the search string, the **Partition** column identifies the partition (file system) where the string is located and the **Seen** column records if the search tool reported at least one instance of the string (yes or no) in meta-data.

Meta-Data Results for Indexed Search of Windows Data Set			
Case	**String**	**Partition**	**Seen**
FT-SS-09-Meta			
	thunderbird	ntfs	Yes
	cañón	fat32	No
	cañón	exfat	No
	cañón	ntfs	Yes

4.1.6 Comments on Indexed Search of Windows Data Set

The following table presents any comments recorded during testing for a test case.

Case	**Comments on Indexed Search of Windows Data Set**
FT-SS-02	Hits on strings "WOLF", "Wolf" and "wolf" were reported twice.
FT-SS-07-CJK-char	The search was run more than once after re indexing and the results were inconsistent. Sometimes all strings were found, other times no strings were found and sometimes only the UTF-8 strings were found.
FT-SS-07-CJK-kana	Building an index was a problem. There were two possible options that could be selected: Japanese or Unicode Multi-lingual Plane. Selecting "Japanese" failed to produce an index and returned a "non-hex char" error. The other option worked once for "みつびし", but then would not yield any hits after re indexing later. No hits were returned for "スバル".
FT-SS-07-Latin	UTF-16 hits were reported twice.
FT-SS-07-RTL	No UTF-16-BE hits were reported.
FT-SS-09-MFT	Also listed in $MFT.
FT-SS-09-Meta	String IDs 2641 & 2645 in NTFS were possibly matched, but the "Search Hits" did not show any context around the string hit. (Therefore, a string ID for the hit was not visible to confirm the reported string.)

4.1.7 Results for Live Search of Windows Data Set

The table columns contain the following information:

- **Case:** The test case identifier.
- **Expected String:** The strings that should be reported by the search.
- **Active Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in an active file.
- **Deleted Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in a deleted file.
- **Unallocated Space:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in unallocated space.
- **Expected:** The number of instances of the expected string found in the group (i.e., Active files, Deleted files or Unallocated space).
- **Hits:** The number of times the expected string was found in the group.
- **Misses:** The number of times the expected string was missed (not found) in the group.

Notes: If the row identifies a test case, then the results are a summary for all the strings that should be found.

In the Expected String column for test case FT-SS-09-DOC each string is labeled to indicate features of the expected string. The labels include the file type (.doc, .docx or .html) and the encoding of the string (if a .doc file). If the string has embedded formatting it is labeled as *Formatted*, e.g., the string *crossbow* has the substring *cross* formatted as bold and underlined, i.e., **cross**bow.

Results for Live Search of Windows Data Set

Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-01		3	3	0	3	3	0	1	1	0
	DireWolf	3	3	0	3	3	0	1	1	0
FT-SS-02		15	15	0	15	15	0	5	5	0
	WOLF	3	3	0	3	3	0	1	1	0
	wolf	3	3	0	3	3	0	1	1	0
	Wolf	3	3	0	3	3	0	1	1	0
	DireWolf	3	3	0	3	3	0	1	1	0
	WereWolf	3	3	0	3	3	0	1	1	0
FT-SS-03		9	9	0	9	9	0	3	3	0
	WOLF	3	3	0	3	3	0	1	1	0

Results for Live Search of Windows Data Set										
Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
	wolf	3	3	0	3	3	0	1	1	0
	Wolf	3	3	0	3	3	0	1	1	0
FT-SS-04		3	3	0	3	3	0	0	0	0
	panda and fox	3	3	0	3	3	0	0	0	0
FT-SS-05		6	6	0	6	6	0	2	2	0
	DireWolf	3	3	0	3	3	0	1	1	0
	WereWolf	3	3	0	3	3	0	1	1	0
FT-SS-06		12	12	0	12	12	0	0	0	0
	fox and not tiger	12	12	0	12	12	0	0	0	0
FT-SS-07-CJK-char		18	18	0	18	18	0	6	6	0
	□ □	9	9	0	9	9	0	3	3	0
	□ □	9	9	0	9	9	0	3	3	0
FT-SS-07-CJK-hangul		9	9	0	9	9	0	3	3	0
	서울	9	9	0	9	9	0	3	3	0
FT-SS-07-CJK-kana		18	18	0	18	18	0	6	6	0
	スバル	9	9	0	9	9	0	3	3	0
	みつびし	9	9	0	9	9	0	3	3	0
FT-SS-07-Cyrillic		9	9	0	9	9	0	3	3	0
	Сибирь	9	9	0	9	9	0	3	3	0
FT-SS-07-Latin		18	18	0	18	18	0	6	6	0
	garçon	9	9	0	9	9	0	3	3	0

Results for Live Search of Windows Data Set

Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
	Schönheit	9	9	0	9	9	0	3	3	0
FT-SS-07-NoBOM		39	39	0	39	39	0	13	13	0
	Россия	9	9	0	9	9	0	3	3	0
	فلافل	9	9	0	9	9	0	3	3	0
	□ □	9	9	0	9	9	0	3	3	0
	QuarterHorse	12	12	0	12	12	0	4	4	0
FT-SS-07-Norm		75	75	0	75	75	0	25	25	0
	mañana (NFD)	9	9	0	9	9	0	3	3	0
	infinity (No Ligature)	12	12	0	12	12	0	4	4	0
	Mäuse (NFD)	9	9	0	9	9	0	3	3	0
	infinity (Ligature)	9	9	0	9	9	0	3	3	0
	Mäuse (NFC)	9	9	0	9	9	0	3	3	0
	libertà (NFC)	9	9	0	9	9	0	3	3	0
	libertà (NFD)	9	9	0	9	9	0	3	3	0
	ma ana (NFC)	9	9	0	9	9	0	3	3	0
FT-SS-07-RTL		9	9	0	9	9	0	3	3	0
	الكسكس	9	9	0	9	9	0	3	3	0
FT-SS-09-Doc		16	16	0	0	0	0	16	13	3
	longbow .html	2	2	0	0	0	0	2	2	0
	shotgun Formatted .doc UTF-16	2	2	0	0	0	0	2	2	0

Results for Live Search of Windows Data Set										
Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
	revolver .doc UTF-16	2	2	0	0	0	0	2	2	0
	peroxide .docx	2	2	0	0	0	0	2	1	1
	nitroglycerin Formatted .docx	2	2	0	0	0	0	2	1	1
	rifle .doc UTF-8	2	2	0	0	0	0	2	2	0
	crossbow Formatted .html	2	2	0	0	0	0	2	1	1
	flintlock Formatted .doc UTF-8	2	2	0	0	0	0	2	2	0
FT-SS-09-Frag		2	2	0	0	0	0	0	0	0
	Washington	1	1	0	0	0	0	0	0	0
	California	1	1	0	0	0	0	0	0	0
FT-SS-09-Lost		0	0	0	0	0	0	4	4	0
	SecretKey	0	0	0	0	0	0	2	2	0
	disconnected	0	0	0	0	0	0	2	2	0
FT-SS-09-MFT		4	4	0	4	4	0	0	0	0
	bear	4	4	0	4	4	0	0	0	0
FT-SS-09-Meta		6	6	0	6	6	0	2	2	0
	cañón	3	3	0	3	3	0	1	1	0
	thunderbird	3	3	0	3	3	0	1	1	0
FT-SS-10-Hex		3	3	0	3	3	0	1	1	0
	panda	3	3	0	3	3	0	1	1	0

Results for Live Search of Windows Data Set										
Case	Expected String	Active Files			Deleted Files			Unallocated Space		
		Expected	Hits	Misses	Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-10-Regex		6	6	0	6	6	0	2	2	0
	DireWolf	3	3	0	3	3	0	1	1	0
	WereWolf	3	3	0	3	3	0	1	1	0

4.1.8 Meta-Data results for Live Search of Windows Data Set

The following table presents search results for strings located in file system meta-data. The **Case** column identifies the test case, the **String** column identifies the search string, the **Partition** column identifies the partition (file system) where the string is located and the **Seen** column records if the search tool reported at least one instance of the string (yes or no) in meta-data.

Meta-Data Results for Live Search of Windows Data Set			
Case	String	Partition	Seen
FT-SS-09-Meta			
	thunderbird	ntfs	Yes
	cañón	fat32	Yes
	cañón	exfat	Yes
	cañón	ntfs	Yes

4.1.9 Comments on Live Search of Windows Data Set

The following table presents any comments recorded during testing for a test case.

Case	Comments on Live Search of Windows Data Set
FT-SS-06	UTF-16 strings are reported twice.
FT-SS-07-Latin	UTF-16 strings are reported twice.
FT-SS-07-NoBOM	UTF-16 strings for "QuarterHorse" are reported twice.
FT-SS-07-Norm	UTF-16 strings normalized as NFC are reported twice.
FT-SS-09-Doc	UTF-16 strings are reported twice.
FT-SS-09-Lost	UTF-16 strings are reported twice.

4.2 Results for Data Set: UNIX

This section provides results for the UNIX data set.

4.2.1 Results for Physical Search of UNIX Data Set

The table columns contain the following information:

- **Case:** The test case identifier.
- **Expected String:** The strings that should be reported by the search.
- **Active Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in an active file.
- **Deleted Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in a deleted file.
- **Unallocated Space:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in unallocated space.
- **Expected:** The number of instances of the expected string found in the group (i.e., Active files, Deleted files or Unallocated space).
- **Hits:** The number of times the expected string was found in the group.
- **Misses:** The number of times the expected string was missed (not found) in the group.

Notes: If the row identifies a test case, then the results are a summary for all the strings that should be found.

In the Expected String column for test case FT-SS-09-DOC each string is labeled to indicate features of the expected string. The labels include the file type (.doc, .docx or .html) and the encoding of the string (if a .doc file). If the string has embedded formatting it is labeled as *Formatted*, e.g., the string *crossbow* has the substring *cross* formatted as bold and underlined, i.e., **<u>cross</u>**bow.

Results for Physical Search of UNIX Data Set

Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-01		4	4	0	4	4	0
	DireWolf	4	4	0	4	4	0
FT-SS-02		20	20	0	20	20	0
	WOLF	4	4	0	4	4	0
	wolf	4	4	0	4	4	0
	Wolf	4	4	0	4	4	0
	DireWolf	4	4	0	4	4	0
	WereWolf	4	4	0	4	4	0
FT-SS-03		12	12	0	12	12	0

Results for Physical Search of UNIX Data Set

Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
	WOLF	4	4	0	4	4	0
	wolf	4	4	0	4	4	0
	Wolf	4	4	0	4	4	0
FT-SS-04		4	4	0	4	4	0
	panda and fox	4	4	0	4	4	0
FT-SS-05		8	8	0	8	8	0
	DireWolf	4	4	0	4	4	0
	WereWolf	4	4	0	4	4	0
FT-SS-06		16	0	16	16	0	16
	fox and not tiger	16	0	16	16	0	16
FT-SS-07-CJK-char		24	24	0	24	24	0
	□ □	12	12	0	12	12	0
	□ □	12	12	0	12	12	0
FT-SS-07-CJK-hangul		12	12	0	12	12	0
	서울	12	12	0	12	12	0
FT-SS-07-CJK-kana		24	24	0	24	24	0
	スバル	12	12	0	12	12	0
	みつびし	12	12	0	12	12	0
FT-SS-07-Cyrillic		12	12	0	12	12	0
	Сибирь	12	12	0	12	12	0
FT-SS-07-Latin		24	24	0	24	24	0
	garçon	12	12	0	12	12	0
	Schönheit	12	12	0	12	12	0
FT-SS-07-NoBOM		52	52	0	52	52	0
	Россия	12	12	0	12	12	0
	فلافل	12	12	0	12	12	0
	□ □	12	12	0	12	12	0
	QuarterHorse	16	16	0	16	16	0

Results for Physical Search of UNIX Data Set							
Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-07-Norm		100	100	0	100	100	0
	mañana (NFD)	12	12	0	12	12	0
	infinity (No Ligature)	16	16	0	16	16	0
	Mäuse (NFD)	12	12	0	12	12	0
	infinity (Ligature)	12	12	0	12	12	0
	Mäuse (NFC)	12	12	0	12	12	0
	libertà (NFC)	12	12	0	12	12	0
	libertà (NFD)	12	12	0	12	12	0
	ma ana (NFC)	12	12	0	12	12	0
FT-SS-07-RTL		12	12	0	12	12	0
	الكسكس	12	12	0	12	12	0
FT-SS-09-Doc		16	13	3	0	0	0
	longbow .html	2	2	0	0	0	0
	shotgun Formatted .doc UTF-16	2	2	0	0	0	0
	revolver .doc UTF-16	2	2	0	0	0	0
	peroxide .docx	2	1	1	0	0	0
	nitroglycerin Formatted .docx	2	1	1	0	0	0
	rifle .doc UTF-8	2	2	0	0	0	0
	crossbow Formatted .html	2	1	1	0	0	0
	flintlock Formatted .doc UTF-8	2	2	0	0	0	0
FT-SS-09-Meta		8	8	0	8	8	0
	cañón	4	4	0	4	4	0
	thunderbird	4	4	0	4	4	0
FT-SS-10-Hex		4	4	0	4	4	0

Results for Physical Search of UNIX Data Set							
Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
	panda	4	4	0	4	4	0
FT-SS-10-Regex		8	8	0	8	8	0
	DireWolf	4	4	0	4	4	0
	WereWolf	4	4	0	4	4	0

4.2.2 Meta-Data results for Physical Search of UNIX Data Set

The following table presents search results for strings located in file system meta-data. The **Case** column identifies the test case, the **String** column identifies the search string, the **Partition** column identifies the partition (file system) where the string is located and the **Seen** column records if the search tool reported at least one instance of the string (yes or no) in meta-data.

Meta-Data Results for Physical Search of UNIX Data Set			
Case	String	Partition	Seen
FT-SS-07-CJK-char			
	□ □	osxj	Yes
	□ □	osxc	Yes
	□ □	apfs	Yes
	□ □	osxj	Yes
	□ □	osxc	Yes
	□ □	apfs	Yes
FT-SS-07-Cyrillic			
	Сибирь	osxj	Yes
	Сибирь	osxc	Yes
	Сибирь	apfs	Yes
FT-SS-07-NoBOM			
	فلافل	osxj	Yes
	فلافل	osxc	Yes
	فلافل	apfs	Yes
	Россия	osxj	Yes

Meta-Data Results for Physical Search of UNIX Data Set			
Case	**String**	**Partition**	**Seen**
	Россия	osxc	Yes
	Россия	apfs	Yes
	□ □	osxj	Yes
	□ □	osxc	Yes
	□ □	apfs	Yes
FT-SS-07-RTL			
	الكسكس	osxj	Yes
	الكسكس	osxc	Yes
	الكسكس	apfs	Yes
FT-SS-09-Meta			
	thunderbird	osxj	Yes
	thunderbird	osxc	Yes
	thunderbird	apfs	Yes
	thunderbird	ext4	Yes
	cañón	ext4	Yes

4.2.3 Comments on Physical Search of UNIX Data Set

The following table presents any comments recorded during testing for a test case.

Case	Comments on Physical Search of UNIX Data Set
FT-SS-06	Tool crashes. page protection fault.
FT-SS-07-Latin	UTF-16 encoded strings are reported twice.
FT-SS-07-NoBOM	UTF-16 strings for "QuarterHorse" are reported twice.
FT-SS-07-Norm	Searches do not use Unicode normalization on the search string. Strings normalized as NFC are reported twice.
FT-SS-09-Doc	UTF-16 strings are reported twice.

4.2.4 Results for Indexed Search of UNIX Data Set

The table columns contain the following information:

- **Case:** The test case identifier.

- **Expected String:** The strings that should be reported by the search.
- **Active Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in an active file.
- **Deleted Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in a deleted file.
- **Unallocated Space:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in unallocated space.
- **Expected:** The number of instances of the expected string found in the group (i.e., Active files, Deleted files or Unallocated space).
- **Hits:** The number of times the expected string was found in the group.
- **Misses:** The number of times the expected string was missed (not found) in the group.

Notes: If the row identifies a test case, then the results are a summary for all the strings that should be found.

In the Expected String column for test case FT-SS-09-DOC each string is labeled to indicate features of the expected string. The labels include the file type (.doc, .docx or .html) and the encoding of the string (if a .doc file). If the string has embedded formatting it is labeled as *Formatted*, e.g., the string *crossbow* has the substring *cross* formatted as bold and underlined, i.e., **cross**bow.

Results for Indexed Search of UNIX Data Set

Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-01		4	4	0	4	4	0
	DireWolf	4	4	0	4	4	0
FT-SS-02		20	20	0	20	20	0
	WOLF	4	4	0	4	4	0
	wolf	4	4	0	4	4	0
	Wolf	4	4	0	4	4	0
	DireWolf	4	4	0	4	4	0
	WereWolf	4	4	0	4	4	0
FT-SS-03		12	12	0	12	12	0
	WOLF	4	4	0	4	4	0
	wolf	4	4	0	4	4	0
	Wolf	4	4	0	4	4	0
FT-SS-04		4	4	0	4	4	0
	panda and fox	4	4	0	4	4	0

Results for Indexed Search of UNIX Data Set							
Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
FT-SS-05		8	8	0	8	8	0
	DireWolf	4	4	0	4	4	0
	WereWolf	4	4	0	4	4	0
FT-SS-06		16	0	16	16	0	16
	fox and not tiger	16	0	16	16	0	16
FT-SS-07-CJK-char		24	8	16	24	8	16
	□ □	12	4	8	12	4	8
	□ □	12	4	8	12	4	8
FT-SS-07-CJK-hangul		12	0	12	12	0	12
	서울	12	0	12	12	0	12
FT-SS-07-CJK-kana		24	0	24	24	0	24
	スバル	12	0	12	12	0	12
	みつびし	12	0	12	12	0	12
FT-SS-07-Cyrillic		12	8	4	12	8	4
	Сибирь	12	8	4	12	8	4
FT-SS-07-Latin		24	20	4	24	20	4
	garçon	12	8	4	12	8	4
	Schönheit	12	12	0	12	12	0
FT-SS-07-NoBOM		52	36	16	52	36	16
	Россия	12	8	4	12	8	4
	فلافل	12	8	4	12	8	4
	□ □	12	4	8	12	4	8
	QuarterHorse	16	16	0	16	16	0
FT-SS-07-Norm		100	48	52	100	48	52
	mañana (NFD)	12	0	12	12	0	12
	libertà (NFD)	12	0	12	12	0	12
	Mäuse (NFD)	12	0	12	12	0	12
	infinity (Ligature)	12	0	12	12	0	12

Results for Indexed Search of UNIX Data Set							
Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
	Mäuse (NFC)	12	12	0	12	12	0
	infinity (No Ligature)	16	16	0	16	16	0
	ma ana (NFC)	12	12	0	12	12	0
	libertà (NFC)	12	8	4	12	8	4
FT-SS-07-RTL		12	8	4	12	8	4
	الكسكس	12	8	4	12	8	4
FT-SS-09-Doc		16	15	1	0	0	0
	longbow .html	2	2	0	0	0	0
	shotgun Formatted .doc UTF-16	2	2	0	0	0	0
	revolver .doc UTF-16	2	2	0	0	0	0
	peroxide .docx	2	2	0	0	0	0
	nitroglycerin Formatted .docx	2	2	0	0	0	0
	rifle .doc UTF-8	2	2	0	0	0	0
	crossbow Formatted .html	2	1	1	0	0	0
	flintlock Formatted .doc UTF-8	2	2	0	0	0	0
FT-SS-09-Meta		8	8	0	8	8	0
	cañón	4	4	0	4	4	0
	thunderbird	4	4	0	4	4	0
FT-SS-10-Hex		4	4	0	4	4	0
	panda	4	4	0	4	4	0

4.2.5 Meta-Data results for Indexed Search of UNIX Data Set

The following table presents search results for strings located in file system meta-data. The **Case** column identifies the test case, the **String** column identifies the search string, the **Partition** column identifies the partition (file system) where the string is located and the **Seen** column records if the search tool reported at least one instance of the string (yes or no) in meta-data.

Meta-Data Results for Indexed Search of UNIX Data Set			
Case	**String**	**Partition**	**Seen**
FT-SS-07-CJK-char			
	□ □	osxj	Yes
	□ □	osxc	Yes
	□ □	apfs	Yes
	□ □	osxj	Yes
	□ □	osxc	Yes
	□ □	apfs	Yes
FT-SS-07-Cyrillic			
	Сибирь	osxj	Yes
	Сибирь	osxc	Yes
	Сибирь	apfs	Yes
FT-SS-07-NoBOM			
	فلافل	osxj	Yes
	فلافل	osxc	Yes
	فلافل	apfs	Yes
	Россия	osxj	Yes
	Россия	osxc	Yes
	Россия	apfs	Yes
	□ □	osxj	Yes
	□ □	osxc	Yes
	□ □	apfs	Yes
FT-SS-07-RTL			
	الكسكس	osxj	Yes
	الكسكس	osxc	Yes
	الكسكس	apfs	Yes
FT-SS-09-Meta			

Meta-Data Results for Indexed Search of UNIX Data Set			
Case	**String**	**Partition**	**Seen**
	thunderbird	osxj	Yes
	thunderbird	osxc	Yes
	thunderbird	apfs	Yes
	thunderbird	ext4	Yes
	cañón	ext4	Yes

4.2.6 Comments on Indexed Search of UNIX Data Set

The following table presents any comments recorded during testing for a test case.

Case	Comments on Indexed Search of UNIX Data Set
FT-SS-06	Tool hangs & crashes.
FT-SS-07-CJK-hangul	Tool crashed.
FT-SS-07-CJK-kana	Failed to index; hung & crash.
FT-SS-07-Cyrillic	No hits on UTF-16-BE.
FT-SS-07-NoBOM	No UTF-16-BE returned for Arabic & Russian. No UTF-16 returned for Traditional Chinese.

4.2.7 Results for Live Search of UNIX Data Set

The table columns contain the following information:

- **Case:** The test case identifier.
- **Expected String:** The strings that should be reported by the search.
- **Active Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in an active file.
- **Deleted Files:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in a deleted file.
- **Unallocated Space:** A group of three columns (**Expected, Hits and Misses**) giving the number of hits and misses when searching for the expected string in unallocated space.
- **Expected:** The number of instances of the expected string found in the group (i.e., Active files, Deleted files or Unallocated space).
- **Hits:** The number of times the expected string was found in the group.
- **Misses:** The number of times the expected string was missed (not found) in the group.

Notes: If the row identifies a test case, then the results are a summary for all the strings that should be found.

In the Expected String column for test case FT-SS-09-DOC each string is labeled to indicate features of the expected string. The labels include the file type (.doc, .docx or .html) and the encoding of the string (if a .doc file). If the string has embedded formatting it is labeled as *Formatted*, e.g., the string *crossbow* has the substring *cross* formatted as bold and underlined, i.e., **cross**bow.

Results for Live Search of UNIX Data Set							
Case	**Expected String**	**Active Files**			**Deleted Files**		
		Expected	**Hits**	**Misses**	**Expected**	**Hits**	**Misses**
FT-SS-01		4	4	0	4	4	0
	DireWolf	4	4	0	4	4	0
FT-SS-02		20	20	0	20	20	0
	WOLF	4	4	0	4	4	0
	wolf	4	4	0	4	4	0
	Wolf	4	4	0	4	4	0
	DireWolf	4	4	0	4	4	0
	WereWolf	4	4	0	4	4	0
FT-SS-03		12	12	0	12	12	0
	WOLF	4	4	0	4	4	0
	wolf	4	4	0	4	4	0
	Wolf	4	4	0	4	4	0
FT-SS-04		4	4	0	4	4	0
	panda and fox	4	4	0	4	4	0
FT-SS-05		8	8	0	8	8	0
	DireWolf	4	4	0	4	4	0
	WereWolf	4	4	0	4	4	0
FT-SS-06		16	16	0	16	16	0
	fox and not tiger	16	16	0	16	16	0
FT-SS-07-CJK-char		24	24	0	24	24	0
	□ □	12	12	0	12	12	0
	□ □	12	12	0	12	12	0
FT-SS-07-CJK-hangul		12	12	0	12	12	0

Results for Live Search of UNIX Data Set							
Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
	서울	12	12	0	12	12	0
FT-SS-07-CJK-kana		24	24	0	24	24	0
	スバル	12	12	0	12	12	0
	みつびし	12	12	0	12	12	0
FT-SS-07-Cyrillic		12	12	0	12	12	0
	Сибирь	12	12	0	12	12	0
FT-SS-07-Latin		24	24	0	24	24	0
	garçon	12	12	0	12	12	0
	Schönheit	12	12	0	12	12	0
FT-SS-07-NoBOM		52	52	0	52	52	0
	Россия	12	12	0	12	12	0
	فلافل	12	12	0	12	12	0
	□ □	12	12	0	12	12	0
	QuarterHorse	16	16	0	16	16	0
FT-SS-07-Norm		100	100	0	100	100	0
	mañana (NFD)	12	12	0	12	12	0
	infinity (No Ligature)	16	16	0	16	16	0
	Mäuse (NFD)	12	12	0	12	12	0
	infinity (Ligature)	12	12	0	12	12	0
	Mäuse (NFC)	12	12	0	12	12	0
	libertà (NFC)	12	12	0	12	12	0
	libertà (NFD)	12	12	0	12	12	0
	ma ana (NFC)	12	12	0	12	12	0
FT-SS-07-RTL		12	12	0	12	12	0
	الكسكس	12	12	0	12	12	0
FT-SS-09-Doc		16	15	1	0	0	0
	longbow .html	2	2	0	0	0	0

Results for Live Search of UNIX Data Set							
Case	Expected String	Active Files			Deleted Files		
		Expected	Hits	Misses	Expected	Hits	Misses
	shotgun Formatted .doc UTF-16	2	2	0	0	0	0
	revolver .doc UTF-16	2	2	0	0	0	0
	peroxide .docx	2	2	0	0	0	0
	nitroglycerin Formatted .docx	2	2	0	0	0	0
	rifle .doc UTF-8	2	2	0	0	0	0
	crossbow Formatted .html	2	1	1	0	0	0
	flintlock Formatted .doc UTF-8	2	2	0	0	0	0
FT-SS-09-Meta		8	8	0	8	8	0
	cañón	4	4	0	4	4	0
	thunderbird	4	4	0	4	4	0
FT-SS-10-Hex		4	4	0	4	4	0
	panda	4	4	0	4	4	0
FT-SS-10-Regex		8	8	0	8	8	0
	DireWolf	4	4	0	4	4	0
	WereWolf	4	4	0	4	4	0

4.2.8 Meta-Data results for Live Search of UNIX Data Set

The following table presents search results for strings located in file system meta-data. The **Case** column identifies the test case, the **String** column identifies the search string, the **Partition** column identifies the partition (file system) where the string is located and the **Seen** column records if the search tool reported at least one instance of the string (yes or no) in meta-data.

Meta-Data Results for Live Search of UNIX Data Set			
Case	String	Partition	Seen
FT-SS-07-CJK-char			

Meta-Data Results for Live Search of UNIX Data Set			
Case	**String**	**Partition**	**Seen**
	□ □	osxj	Yes
	□ □	osxc	Yes
	□ □	apfs	Yes
	□ □	osxj	Yes
	□ □	osxc	Yes
	□ □	apfs	Yes
FT-SS-07-Cyrillic			
	Сибирь	osxj	Yes
	Сибирь	osxc	Yes
	Сибирь	apfs	Yes
FT-SS-07-NoBOM			
	فلافل	osxj	Yes
	فلافل	osxc	Yes
	فلافل	apfs	Yes
	Россия	osxj	Yes
	Россия	osxc	Yes
	Россия	apfs	Yes
	□ □	osxj	Yes
	□ □	osxc	Yes
	□ □	apfs	Yes
FT-SS-07-RTL			
	الكسكس	osxj	Yes
	الكسكس	osxc	Yes
	الكسكس	apfs	Yes
FT-SS-09-Meta			
	thunderbird	osxj	Yes
	thunderbird	osxc	Yes
	thunderbird	apfs	Yes

Meta-Data Results for Live Search of UNIX Data Set			
Case	String	Partition	Seen
	thunderbird	ext4	Yes
	cañón	ext4	Yes

4.2.9 Comments on Live Search of UNIX Data Set

The following table presents any comments recorded during testing for a test case.

Case	Comments on Live Search of UNIX Data Set
FT-SS-07-Latin	Hits on strings encoded UTF-16 are reported twice.
FT-SS-07-NoBOM	UTF-16 hits for "QuarterHorse" are reported twice.
FT-SS-07-Norm	No Unicode normalization. Strings normalized as NFC and without a ligature encoded as UTF-16 are reported twice. Some strings are not found unless searched for alone without any other strings included in the search.

4.3 Unicode Normalization

The following is from "Unicode® Standard Annex #15, Unicode Normalization Forms." http://unicode.org/reports/tr15/

> Unicode Normalization Forms are formally defined normalizations of Unicode strings which make it possible to determine whether any two Unicode strings are equivalent to each other. Depending on the particular Unicode Normalization Form, that equivalence can either be a canonical equivalence or a compatibility equivalence.
>
> Essentially, the Unicode Normalization Algorithm puts all combining marks in a specified order, and uses rules for decomposition and composition to transform each string into one of the Unicode Normalization Forms. A binary comparison of the transformed strings will then determine equivalence.
>
> The four Unicode Normalization Forms are summarized in *Table 1.*
>
> Table 1. Normalization Forms

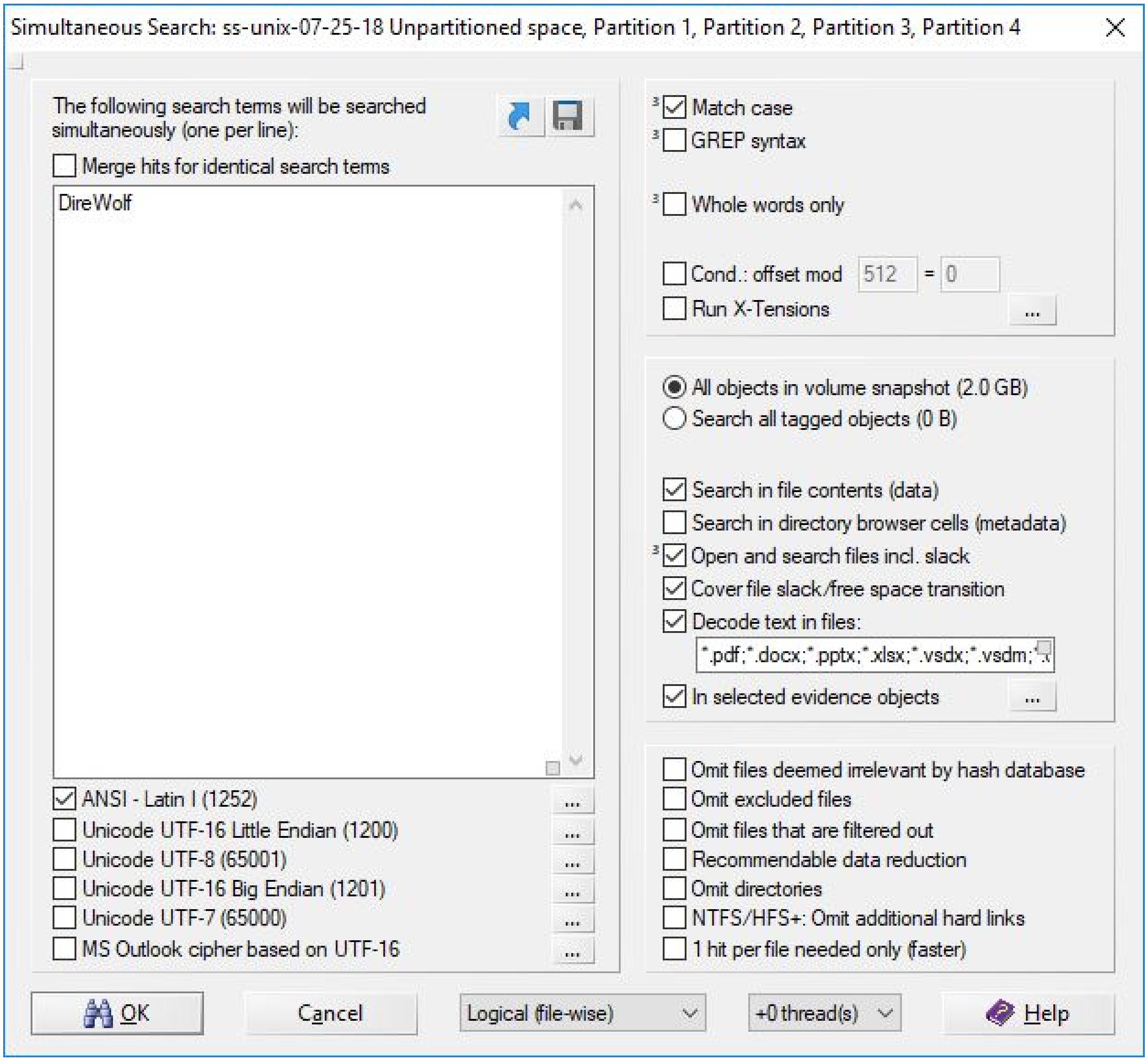
Form	Description
Normalization Form D (NFD)	Canonical Decomposition
Normalization Form C (NFC)	Canonical Decomposition, followed by Canonical Composition
Normalization Form KD (NFKD)	Compatibility Decomposition
Normalization Form KC (NFKC)	Compatibility Decomposition, followed by Canonical Composition

5 Search Configuration Settings

This section presents screen captures illustrating search parameter selection for test searches.

5.1 Logical Search Configuration

The following screen capture is for test case FT-SS-01. The search string is “DireWolf,” with option “Match case” selected, “Whole word” is not selected, “Unicode” is not selected and “ASCII” is selected. The search engine is “Logical” (Live).



Simultaneous Search: ss-unix-07-25-18 Unpartitioned space, Partition 1, Partition 2, Partition 3, Partition 4

The following search terms will be searched simultaneously (one per line):

□ Merge hits for identical search terms

```
DireWolf
```

☑ ANSI - Latin I (1252)
□ Unicode UTF-16 Little Endian (1200)
□ Unicode UTF-8 (65001)
□ Unicode UTF-16 Big Endian (1201)
□ Unicode UTF-7 (65000)
□ MS Outlook cipher based on UTF-16

[3] ☑ Match case
[3] □ GREP syntax
[3] □ Whole words only
□ Cond.: offset mod 512 = 0
□ Run X-Tensions

● All objects in volume snapshot (2.0 GB)
○ Search all tagged objects (0 B)

☑ Search in file contents (data)
□ Search in directory browser cells (metadata)
[3] ☑ Open and search files incl. slack
☑ Cover file slack/free space transition
☑ Decode text in files: *.pdf;*.docx;*.pptx;*.xlsx;*.vsdx;*.vsdm;*.
☑ In selected evidence objects

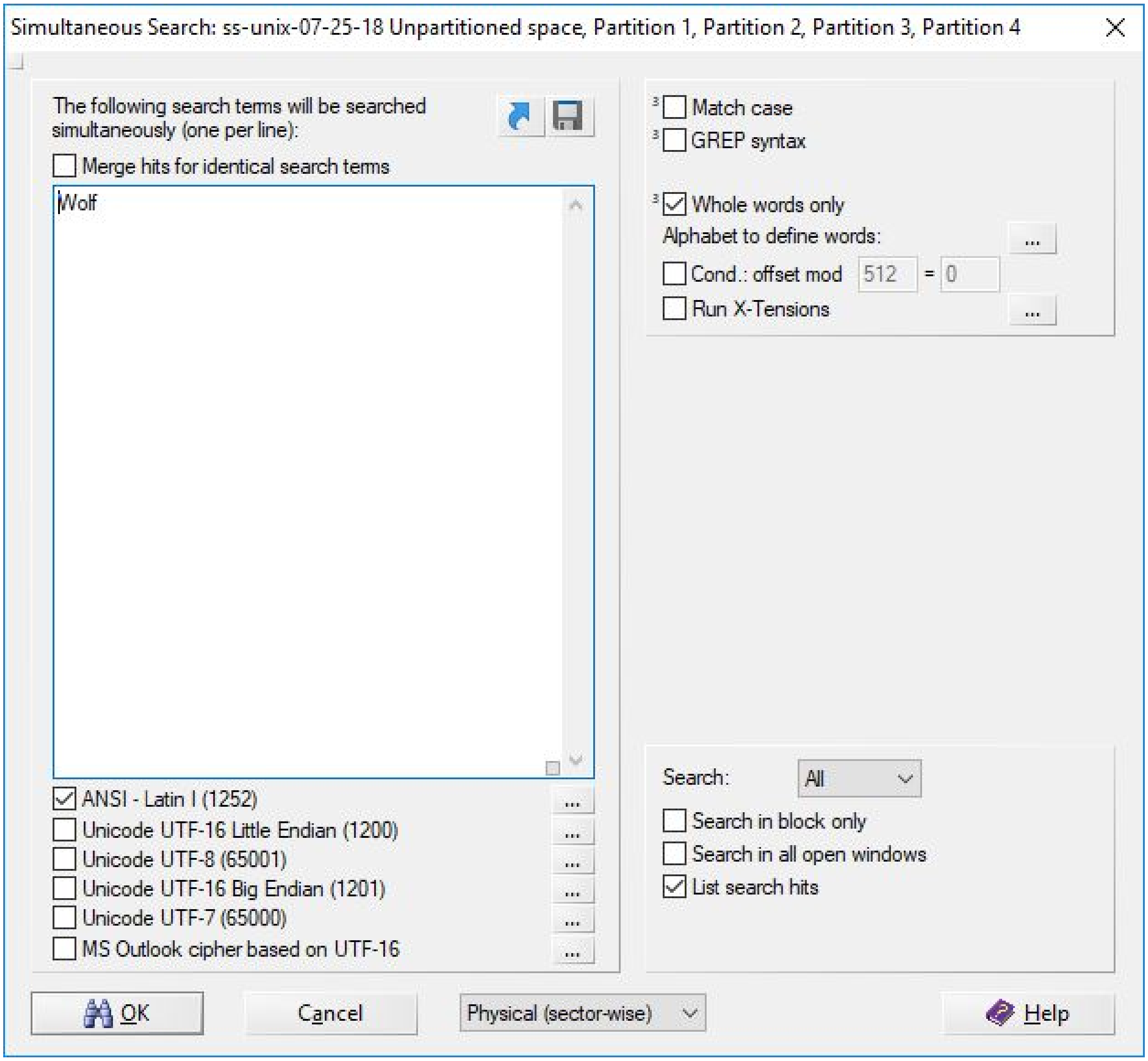
□ Omit files deemed irrelevant by hash database
□ Omit excluded files
□ Omit files that are filtered out
□ Recommendable data reduction
□ Omit directories
□ NTFS/HFS+: Omit additional hard links
□ 1 hit per file needed only (faster)

Logical (file-wise)
+0 thread(s)

Figure 1 Simultaneous (Live) Search Configuration

5.2 Physical Search Configuration

The following screen capture is for test case FT-SS-03. The search string is “Wolf,” “Match case” is not selected, “Whole word” is selected, “Unicode” is not selected and “ASCII” is selected. The search engine is “Physical.”



Simultaneous Search: ss-unix-07-25-18 Unpartitioned space, Partition 1, Partition 2, Partition 3, Partition 4

The following search terms will be searched simultaneously (one per line):

□ Merge hits for identical search terms

```
Wolf
```

☑ ANSI - Latin I (1252)
□ Unicode UTF-16 Little Endian (1200)
□ Unicode UTF-8 (65001)
□ Unicode UTF-16 Big Endian (1201)
□ Unicode UTF-7 (65000)
□ MS Outlook cipher based on UTF-16

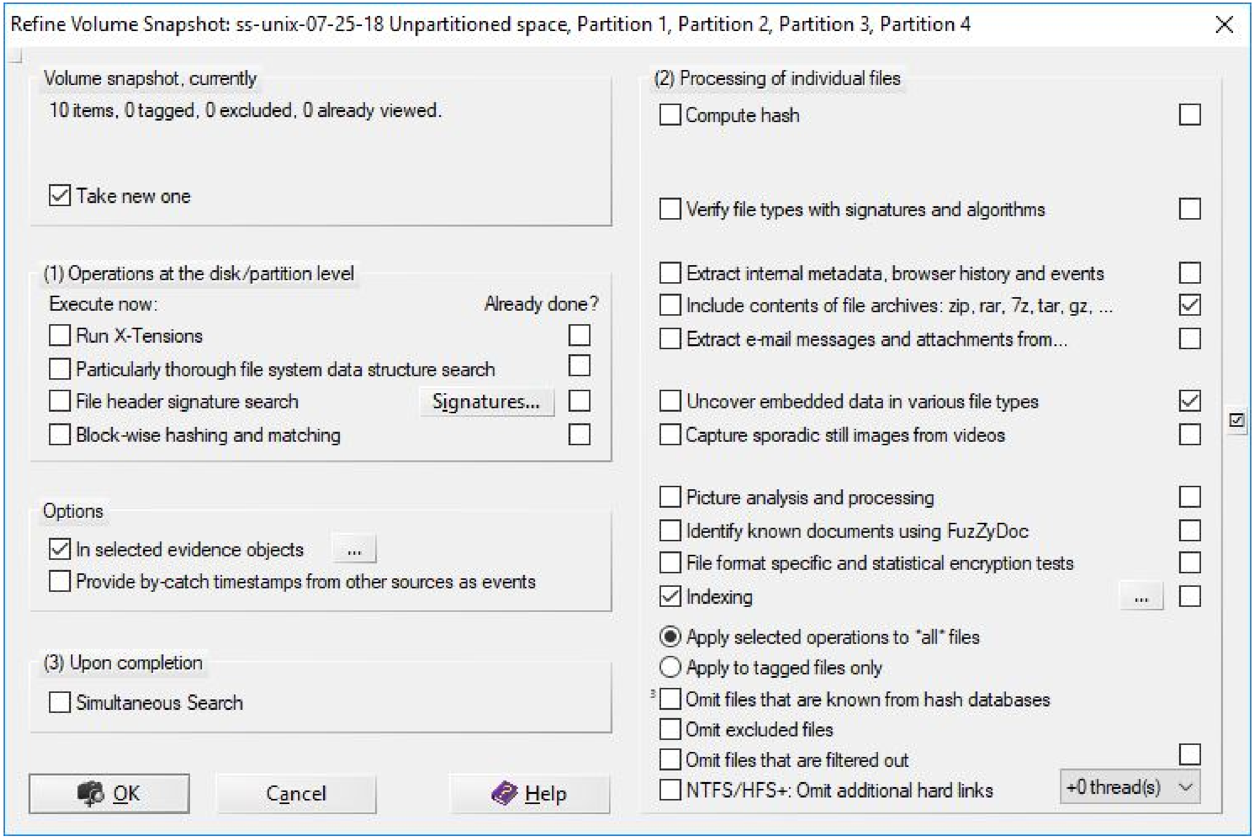
□ Match case
□ GREP syntax
☑ Whole words only
Alphabet to define words:
□ Cond.: offset mod 512 = 0
□ Run X-Tensions

Search: All
□ Search in block only
□ Search in all open windows
☑ List search hits

Figure 2 Physical Search Configuration

5.3 Indexed Search

This screen captures in this section are from test case FT-SS-07-Cyrillic. An index must be built before any searching can be done. The first screen is for the “Refine case” option. The second screen is for building the index and the last screen is for searching with the index.



Refine Volume Snapshot: ss-unix-07-25-18 Unpartitioned space, Partition 1, Partition 2, Partition 3, Partition 4

Volume snapshot, currently
10 items, 0 tagged, 0 excluded, 0 already viewed.
☑ Take new one

(1) Operations at the disk/partition level
Execute now: Already done?
□ Run X-Tensions □
□ Particularly thorough file system data structure search □
□ File header signature search Signatures... □
□ Block-wise hashing and matching □

Options
☑ In selected evidence objects ...
□ Provide by-catch timestamps from other sources as events

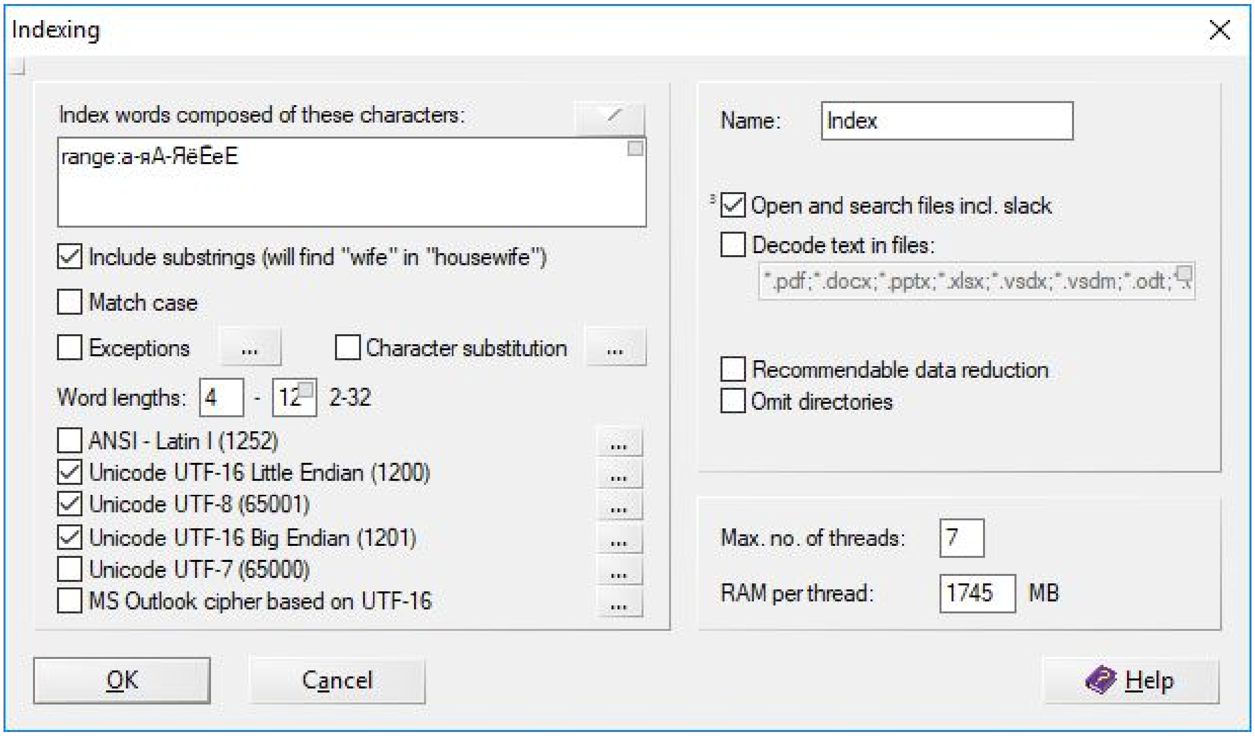
(3) Upon completion
□ Simultaneous Search

OK Cancel Help

(2) Processing of individual files
□ Compute hash □
□ Verify file types with signatures and algorithms □
□ Extract internal metadata, browser history and events □
□ Include contents of file archives: zip, rar, 7z, tar, gz, ... ☑
□ Extract e-mail messages and attachments from... □
□ Uncover embedded data in various file types ☑
□ Capture sporadic still images from videos □
□ Picture analysis and processing □
□ Identify known documents using FuzZyDoc □
□ File format specific and statistical encryption tests □
☑ Indexing ... □
● Apply selected operations to "all" files
○ Apply to tagged files only
□ Omit files that are known from hash databases
□ Omit excluded files
□ Omit files that are filtered out □
□ NTFS/HFS+: Omit additional hard links +0 thread(s)

Figure 3 Refine Case -- Setup for indexing

This index is for Russian language strings in Unicode, UTF-16 and UTF-8.



Indexing

Index words composed of these characters:

range:а-яА-ЯёЁеЕ

☑ Include substrings (will find "wife" in "housewife")

□ Match case

□ Exceptions ... □ Character substitution ...

Word lengths: 4 - 12 2-32

□ ANSI - Latin I (1252) ...

☑ Unicode UTF-16 Little Endian (1200) ...

☑ Unicode UTF-8 (65001) ...

☑ Unicode UTF-16 Big Endian (1201) ...

□ Unicode UTF-7 (65000) ...

□ MS Outlook cipher based on UTF-16 ...

Name: Index

☑ Open and search files incl. slack

□ Decode text in files:

.pdf;.docx;*.pptx;*.xlsx;*.vsdx;*.vsdm;*.odt;*.

□ Recommendable data reduction

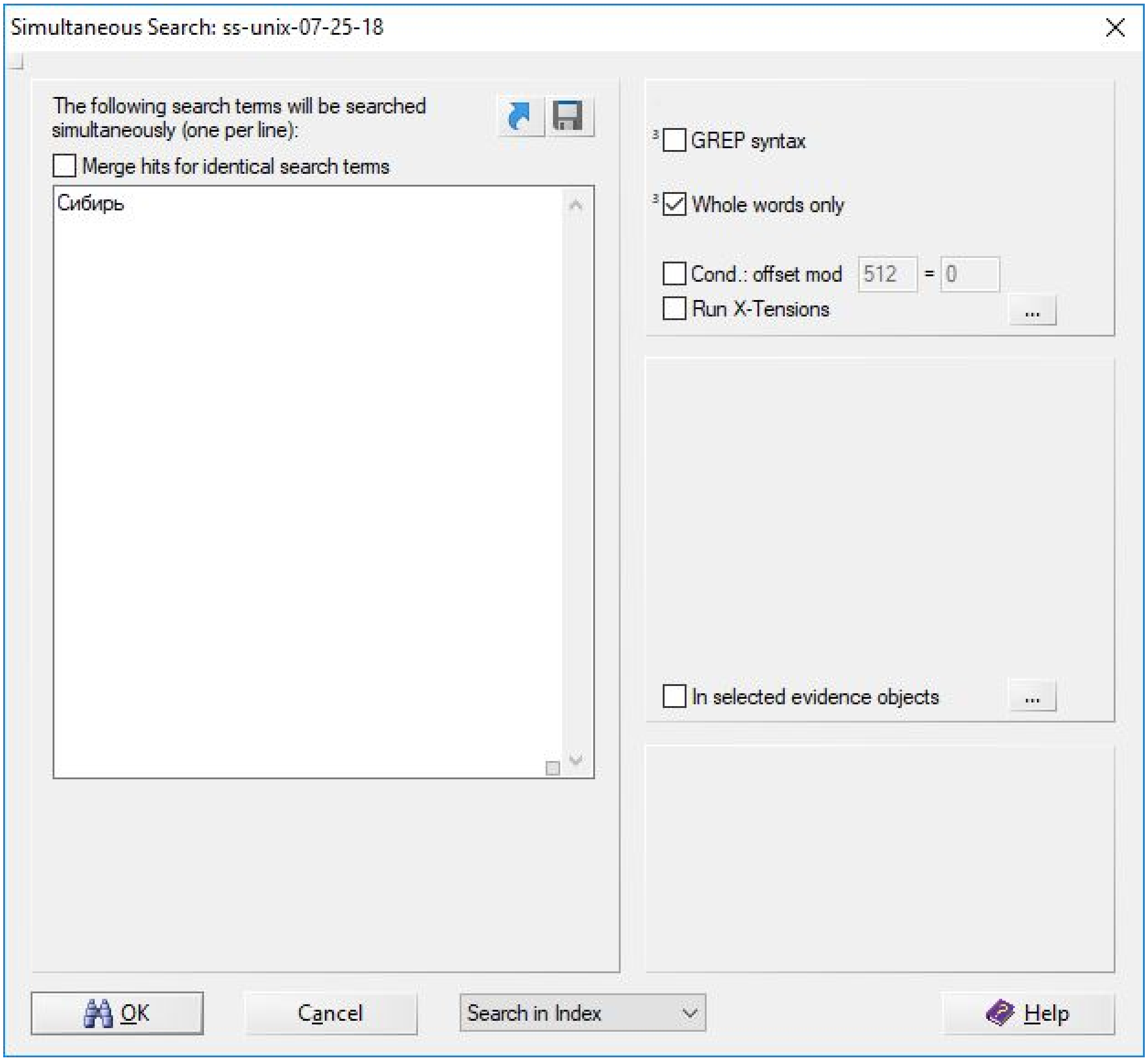
□ Omit directories

Max. no. of threads: 7

RAM per thread: 1745 MB

Figure 4 Build an Index for Russian Language Text

The last screen is for executing the search for “сибирь” (Siberia).



Simultaneous Search: ss-unix-07-25-18

The following search terms will be searched simultaneously (one per line):

□ Merge hits for identical search terms

Сибирь

³□ GREP syntax

³☑ Whole words only

□ Cond.: offset mod 512 = 0

□ Run X-Tensions

□ In selected evidence objects

Search in Index

Figure 5 Do Indexed Search for String Сибирь

END of REPORT