

Odyssey Digital Forensics™

Who uses Odyssey Digital Forensics

Investigators searching for key terms and names of people and places on acquired disk images.

Why they use it

A word may be written in any of dozens of different ways, based on the language, encoding format, and operating system. Without Odyssey, analysts must know which variations to look for and can easily specify the wrong ones — *missing valid matches and wasting time*.

The Odyssey difference

Odyssey recognizes all valid keyword instances regardless of variations so analysts need only submit the word once.

For more information on Odyssey Digital Forensics or for pricing details contact us at (617) 386-2090 or info2007@basistech.com

Digital Forensics advanced keyword search

There are many ways to store a single word on a hard drive — using different code pages, Unicode formats, and file formats. There are also many ways to encode characters with diacritics (accents). For example, the character **á** can be stored as one Unicode character or two (one for the letter and one for the accent).

Here's the problem: Conventional tools require the forensic investigator to manually specify all variations. That wastes time and risks missing valid matches.

Odyssey Digital Forensics™ is software that finds all keyword variations with one search. Odyssey combines industry-leading language technology from Basis Technology — the Rosette® Linguistics Platform — with a high-performance search system that can analyze disk image files acquired from standard forensic tools.

With one query, Odyssey Digital Forensics finds many keyword variations. Unlike other forensic tools, Odyssey doesn't simply return embedded binary matches (potential false positives) that happen to correspond to a particular byte code. Odyssey recognizes text regardless of whether the text is:

- Displayed left to right or right to left (as in Middle-Eastern languages)
- Stored with bits aligned left to right or right to left ("little Endian" or "big Endian")
- Encoded in UTF-8, UTF-16, or UTF-32 Unicode or any of dozens of legacy text encoding systems

Odyssey is the only tool that takes all these variations into account. It also resolves language-specific variations (like those in Table 1). For example, vocalization marks in Arabic are often absent and different combinations of vocalizations may be stored as different sets of Unicode or as different code page characters.

services

You Want Normal Text

Even though you may receive text in multiple language-specific variations and encoding formats, you don't have to conduct searches that way, or save results that way. Odyssey performs several steps to normalize text for ease-of-use:

- Detects the language and encoding of each document
- Converts all text to one encoding (Unicode)
- Normalizes European characters with diacritics to one or two characters
- Adds vowel markings (vocalizations) to Arabic text — normally written without vowels — to reduce ambiguity
- Normalizes the number of Arabic “kashida” — a stylistic calligraphic flourish in Arabic and Japanese

Odyssey Specifications

Odyssey runs on	Windows XP
Odyssey supports these file systems	FAT12 FAT16 FAT32 NTFS EXT2/3 UFS1/2
Odyssey supports these image formats	AFF Encase Raw

Basis Technology Corporation

T 1.617.386.2000
1.800.697.2062 (toll-free)
F 617.386.2020
E info@basistech.com

Four Examples of language-specific variations

An analyst should not have to know all the variations that might apply to a language. Odyssey automatically looks for all variations when conducting a search.

Look-alike characters	العربي Alef maksura (U+0649)	The alef maksura, yeh, and farsi yeh are examples of characters with different Unicode code points, but which look similar and are used in different languages. (The character that changes is the one on the far left of each word.)
	العربي Yeh (U+064A)	
	العربي Farsi yeh (U+06CC)	
Arabic numerals	١٥٤٦ == ١٥٤٦ == 1546	
Full / half-width Japanese	Full-width	Half-width
	カタカナ	かたかな
Traditional vs. Simplified Chinese	Simplified	Traditional
	计算机	電腦

Master Foreign Languages

Using Odyssey Digital Forensics non-native speakers can quickly determine which foreign text merits further investigation and which text does not.

Odyssey uses **Named Entity Extraction** technology to identify words that represent the names of locations, people, and dates. It then uses **Transliteration** technology to convert the named entities from their original non-Roman script to a standardized Romanized form. Take for example the Arabic name وادي أبو شهامة which is then translated to the Romanized form of “Wadi Abu-Shahamah”. This enables a non-native speaker to quickly determine the people, places, and dates contained in a document.

To perform a search, users simply type or copy and paste keywords into the search field — or submit a file of keywords. Odyssey returns the names of files (including deleted files) that contain matches. Click on a file to display the keywords highlighted within surrounding text. Click on “next” to review keyword matches in sequence.

Results are returned and saved in normalized form — a consistent orthographic style and encoding format — which makes handling foreign text much easier (see sidebar). The entire original text is also saved.

Other capabilities include:

- Boolean search
- Bookmarking
- Search history
- Integrity checks to assure the acquired image has not been altered