Digital Forensics: A Primer on Keyword Searching

A typical digital forensic investigation can consist of reviewing and searching through hundreds of thousands of files on a computer system, mobile phone, or tablet. Not every file is going to be of interest or pertinent to the matter. The cost of reviewing files is always a factor and can add up very quickly when billing at a high hourly rate.

This is where keyword searching is one option that can be used to sift through the data set. With the right set of keywords, an examiner can filter down a large data set to locate potentially relevant information quickly. However, there are some common mistakes that attorneys often make when providing search terms that can have the opposite effect, costing more money and returning a large data set full of irrelevant or false positive search hits.

Here are some recommendations to consider when developing a list of keywords for your expert to use:

What types of keyword searching can be performed across the data set?

A list of keywords can take many forms, and the way they are formatted does matter. There are a variety of programs available to a digital forensic examiner capable of performing keyword searches but, not all programs are the same. Some may run searches differently or support other search methodologies. As an example, some programs will support types of Boolean searching and others may not.

Therefore, it is important to consult your expert before developing the list of keywords so that the format of the terms and capabilities are understood. Working together with your expert to generate a list of acceptable keywords will save both time and money.

Bad keywords - what are they?

Some keywords return too many results which will almost certainly not be relevant. One of the most common examples is using short keywords or

acronyms. Terms such as “NDA,” “OR,” or the initials of a person will often return large amounts of irrelevant results. Short terms can occur within a data set thousands of times, even embedded within the programming language of system and program files.

Using terms related to the type of device that is being searched is not a good idea. For example, running a search for “Apple” across a MacBook, iPhone, or iPad is going to turn up an extremely large number of irrelevant results. The same can be said for running the terms “Microsoft” or “window” across a Windows operating system computer.

Using a person’s name as a search term is often not as helpful as you might think, especially if only using their first name. If the search is being performed on their device there is bound to be a vast quantity of irrelevant search hits across the device. The username may appear within the metadata of every document ever created on the system. It certainly will appear in every email message. The hits can likely occur within system artifacts that will have very little relevance to most investigations.

Think about it this way – the user, “Robert” probably has a user account named Robert, and the system’s folder structure and underlying operating system architecture will almost certainly have many references to that keyword. On mobile devices such as tablets or cell phones, there can be multiple contacts with the same first name.

Search terms that contain special characters can often be problematic. Remember the discussion above about different programs having different searching capabilities? Often, special characters, such as star *****, quotation marks **“”**, slashes **/ **, parenthesis **()**, and the ampersand **&** are used as search operators. Terms that contain those characters can obstruct the searching or even return unintended results.

What can be done to help improve keyword searching?

The best thing that can be done to improve the efficiency of keyword searching is to consult with your expert and work with them throughout the process. They can give you insight into what types of keyword searches are

supported by the tools and software that are being used to conduct the analysis. They can also provide suggestions as to what keywords are bad search terms and how to come up with good keywords. At the end of the day, discussing and developing effective keyword searches with your expert will save you and your client both time and money.