The Automatic Extraction of Digital Evidence from Diverse Encoding Files by COM Technique

Wu, Kuo-Ching
Department of Information Management, Central Police University,
Kuei-Shan ,Tao-Tuan, Taiwan
wkc@mail.cpu.edu.tw

Abstract

The component object model (COM) is a quite mature technique, which is used widely in information technology (IT) application, for example, a network distributed (especially cross the Internet) process or an operation of client-server architecture. Nowadays, there are many computer languages, which provide a variety of libraries via COM technique for program designers. The designers may use versatile of computer languages in different operating systems to call interoperable functions encoded with COM. This characteristic is also in the EnScript language provided by EnCase software. It can enable investigators or examiners of digital evidence to design some customized programs for automatically building a mutual communication with other computer languages via COM, and accomplishing some functions that are not provided by or very clumsy in the EnCase software. For example, in EnCase Forensic, users may present some charts by Excel, but they may not find the direct way to do it by themselves, Hence, user can still use the indirect way to fulfill this chart function in EnCase.

Until now, in the commercial and academic fields, it is very rare to discuss or provide the related source code for extracting evidential cue from diverse encoding systems by EnCase EnScript scripting language and COM technology, for example, BIG5, UFT-8, Unicode little/big endian, and so on. Thus, in this paper, we will discuss about automatic extraction and analysis of digital evidence under different encoding system of files, understand how to design EnScript programs and interoperate between EnCase and Microsoft Office Word/Excel environments. In order to fulfill this purpose, in this research it will be conducted by an experimental method to present the results including the deleted or deleted overwritten file of live acquisition. In the final part of this paper we will also give the suggestion remarks and some applications by EnScript language in the future.

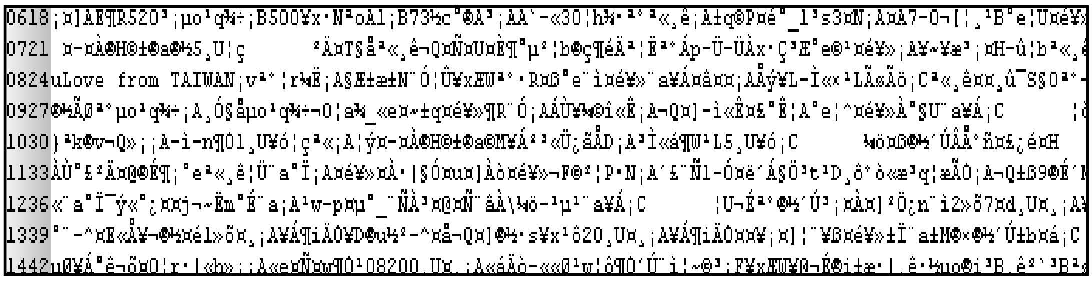
Keywords: Digital Evidence, Component Object Model (COM), Extraction, EnScript Scripting Language, Digital Forensics.

1. Introduction

Until now, the task for the traditional examination of physical evidence may not be able to automate completely since it is not a digital material. Hence, it is difficult to save considerable time or avoid interfering in it. The evidence examined in digital forensics is mostly in storage media, therefore, it is easy to implement automatic procedures in digital forensics.

There are several factors we need to pay special attention: (1) the digital crime materials are often existed in any suspicious computer system, cell phone, PDA, or any handheld device in criminal place; (2) the file size may be too large to manipulate; (3) its content is easily destroyed or contaminated; Hence, it is very important to maintain the chain of custody for digital evidence so that it can be validated in court, and restrictly accessed to any crime scene, lab and computer system.

Investigators or examiners must demonstrate the reasonable suspected evidence in court and how criminal activity was carried out in the case related to cyber crime. Both data acquisition and extraction may be the most challenging task for digital forensics tools. Data acquisition formats are usually different in raw data and may be vendor-specific proprietary. The functions of data viewing and keyword look-up of extraction implemented by these forensics tools may be hard to operate due to the conflict or inconsistency of different encoding systems, file types through the Internet, or internal code in different computer systems, as shown in Figure 1. Besides the problems of encoding and types, data analysis implemented by forensics tools, especially looking for the evidence via keywords look-up function, may be a time-consuming task for investigators or examiners. Hence, it is necessary to design specific task-oriented programs for extracting or retrieving data from evidential files.



```
0618¡¤]ÀÆ¶R520³¡µo¹q¾÷¡B500¥x·N²oÀl¡B73½c°®À³¡AA`-«30¦h¾·²°²«¸ê¡A±q®P¤é°_l³s3¤N¡A¤A7-O¬[¦¸¹B°e¦U¤é¥»
0721 ¤-¤À®H®±®a®½5¸U¦ç      ²Ä¤T§å²«¸ê¬Q¤Ñ¤U¤È¶°µ²¦b®ç¶éÄ²¦Ë²°Áp-Ü-ÜÀx·Ç³Æ°e®¹¤é¥»¡A¥~¥æ³¡¤H-û¦b²«¸ê
0824uLove from TAIWAN¡v²°¦r¼Ë¡A§Æ±æ±N¨Ó¦Û¥xÆW²°·R¤ß°e¨ì¤é¥»¨a¥Á¤â¤¤¡AÅý¥L-Ì«x¹LÃøÃö¡C²«¸ê¤¤¸û¯S§O²°-
0927®½ÃØ²°µo¹q¾÷¡A¸Ó§åµo¹q¾÷¬O¦a¾_«e¤~±q¤é¥»¶R¨Ó¡AÁÙ¥¼®ì«Ê¡A¬Q¤]-ì«Ê¤£°Ê¦A°e¦^¤é¥»À°§U¨a¥Á¡C      ¦
1030}²k®v¬Q»¡¡A-ì-n¶Ò1¸U¥ó¦ç²«¡A¦ý¤-¤À®H®±®a©M¥Á²³«Ü¿ãÅD¡A³Ì«á¶W¹L5¸U¥ó¡C      ¾ö¤ß®½´ÚÅÅ°ñ¤£¿é¤H
1133ÀÜ°£²Ä¤@®É¶¡°e²«¸ê¦Ü¨a°Ï¡A¤é¥»¤À·|§Ó¤u¤]Àò¤é¥»¬F®²¦P·N¡A´£¨Ñ1-Ó¤ë´Á§Ö³t¹D¸ô°ò«æ³q¦æÃÒ¡A¬Q±ß9®É´Ñ
1236«¨a°Ï¯ý«°¿¤¤j¬~Ëm°É¨a¡A¹w-p¤µ°_¨ÑÀ³¤@¤Ñ¨âÀ\¾ö-¹µ¹¨a¥Á¡C      ¦U¬É²°®½´Ú³¡¤À¤]²Ö¿n¨ì2»õ7¤d¸U¤¸¡A¥
1339°¨-^¤Æ«Å¥¬®½¤él»õ¤¸¡A¥Á¶iÄÒ¥D®u½²-^¤å¬Q¤]®½·s¥x¹ô20¸U¤¸¡A¥Á¶iÄÒ¤¤¥¡¤]¦¨¥ß¤é¥»±Ï¨a±M®×®½´Ú±b¤á¡C
1442u0¥Á°ê¬õ¤Q¦r·|«b»¡¡A«e¤Ñ¤w¶Ò¹08200 U¤ ¡A«áÄò-««0¹w¦ô¶Ô´Ú¨ì¦~®³¡F¥xÆW¥0¬É®i±æ·|¸ê·¼uo®i³B¸ê²`³B²
```

Figure 1. unrecognized internal encoding systems.

In this paper, we will focus on discussing the conflict or inconsistency of different encoding systems (such as Big-5, UTF-8, Unicode, …) between EnCase Forensic software and other application software, and how to influence on data viewing and keyword search hits, which is implemented by EnCase search function. To solve this problem, we use the feature of file conversion supported by Microsoft Word, develop programs with the EnScript language and VBA macro language supported by Microsoft Word and Excel, and execute for automatic extraction of digital evidence via the technique of component object model (COM). In order to prove and meet this purpose, the experimental samples of text files from different encoding systems which including ANSI (Big-5), Unicode little endian, Unicode big endian, and UTF-8, as well as other file types, including txt, data, csv, mht, mht, doc(x), and xls(x), are all under consideration. Finally, we will suggest that there are some flaws in the EnScript language.

2. Overview

2.1 Concept of Component Object Model

Component Object Model (COM), a software standard of binary (executable or machine) code and also an open commercial standard, provides a model standard for component interoperability, independent on any particular programming language and multiple platforms (hardware and operating system combination). COM, supported by Microsoft or other vendors, is used to many applications, such as controls, compound documents, automation, data transfer, storage and naming, and so on.

COM employs a standard mechanism to construct virtual function tables (vtables) in main memory and execute call functions of interoperability through the vtables and pointers. Many computer languages with vtable construction may be used to write components and interoperate with other components of the same binary standard.

COM technology enables software components to communicate with each other, for example, Microsoft COM allows Word documents to dynamically link to data (cells) in Excel spreadsheets. COM automation allows programmers to lay out any script in their applications to perform routine tasks or control one application from another. COM provides mechanism for communications among components, even cross processes and network boundaries, shared memory management between components, error and status reporting, and dynamic loading of components. In Figure 2, the component software architecture of COM defines a binary standard for component interoperability, language-independent and provided on multiple platforms [10].

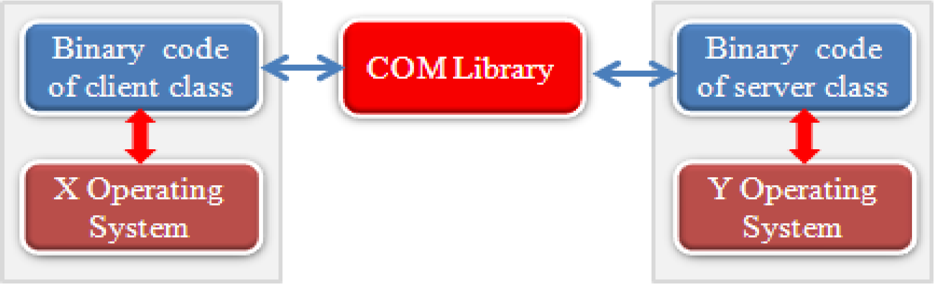


Figure 2. the component interoperability between client and server.

2.2 Digital Evidence

Digital evidence or electronic evidence is any probative information stored or transmitted in digital form that a party of a court case may use at trial [7]. Evidence with digital information (including any data or program) can prove that a crime has been committed, can provide a link between a crime and its victim, or can provide a link between a crime and its perpetrator [4]. The information in memory, on the hard disk, or in any handheld devices are sources of digital evidence.

From the point of criminal law of Taiwan, the definition of digital evidence is the one stored in electromagnetic records by the form of words, sounds, pictures, images, symbols, or others by custom or special regulation, of which the content contains any meaning or intention towards a violation of the laws. Here, the evidence must be proven and demonstrated by specific ways of machines and presented the fact related to cyber crime. The term "evidence" implies the fact that the collection of material must be recognized by forensics and court.

From the perspective of Locard exchange principle — "with contact between two items, there will be an exchange," although the perpetrator(s) of a crime comes into contact with the scene, so the perpetrator(s) will both bring something into the scene and leave with something from the scene [8], digital evidence is easily volatile, damaged or contaminated at a crime scene or criminal laboratory, hence digital evidence must be handled carefully to preserve the integrity of the physical device as well as the information it contains. Some digital evidence requires special collection, packaging, and transportation techniques. Data can be damaged or altered by electromagnetic fields such as those generated by static electricity, magnets, radio transmitters, and other devices. Communication devices such as mobile phones, smart phones, PDAs, and pagers should be secured and prevented from receiving or transmitting messages (packets) once they are identified and collected as evidence.

Information (data and program) collection and recovery which stored on any physical device or transmission (ex., Internet) are very important to the integrity of digital evidence. Persistent information may be stored on some media storages which can be persistently preserved even though the computer (power) is turned off. Volatile information may be stored in memory, or exists in transit (ex., electronic signal or packet of network communication), that will be lost or damage when the computer power is below normal voltage, failure or turned off. Volatile information or file archive for personal computer may temporarily reside in register or cache of CPU, random access memory (RAM) of memory, pagefile.sys provided for virtual memory, not yet overwritten or partial overwritten deleted file stored in disk or USB, and so on.

2.3 Evidence Acquisition and Extraction

All computer forensics tools, both hardware and software, perform specific functions. Table 1 is an example of comparison among forensics vendor’s tools. These functions are grouped into five major categories (forensic phases of SOPs), each with subfunctions for further refining data analysis and recovery: [1]

(1) Acquisition;
(2) Validation and discrimination;
(3) Extraction;
(4) Reconstruction;
(5) Reporting.

Table 1. Comparison of forensics tool functions [1].

Products Functions	ProDiscover Investigator	AccessData Ultimate Toolkit	Guidance Software EnCase
I. Acquisition			
Physical copy and data copy	✓	✓	✓
Data acquisition formats	✓	✓	✓
Remote acquisition	✓		✓*
Ⅱ. Validation and discrimination			
Hashing	✓	✓**	✓**
Filtering	✓	✓	✓
Analyzing file headers	✓	✓	✓
Ⅲ. Extraction			
Data viewing	✓	✓***	✓***
Keyword searching	✓	✓	✓
Decompressing		✓	✓
Carving	✓	✓	✓
Ⅵ. Reconstruction			
Disk-to-disk copy	✓	✓	✓
Image-to-disk copy	✓	✓	✓
Ⅴ. Reporting			
Log report	✓	✓	✓
Report generator	✓	✓	
Ⅵ.Automation features			
Scripting language	✓		✓
Background processing			✓

*Must purchase EnCase Enterprise Edition for this feature.

**Both MD5 and SHA-1 hashing are available.

***Supported file formats vary. EnCase provides for Unicode little and big endian, UTF-8 in traditional Chinese characters, not yet supported in AccessData FTK.

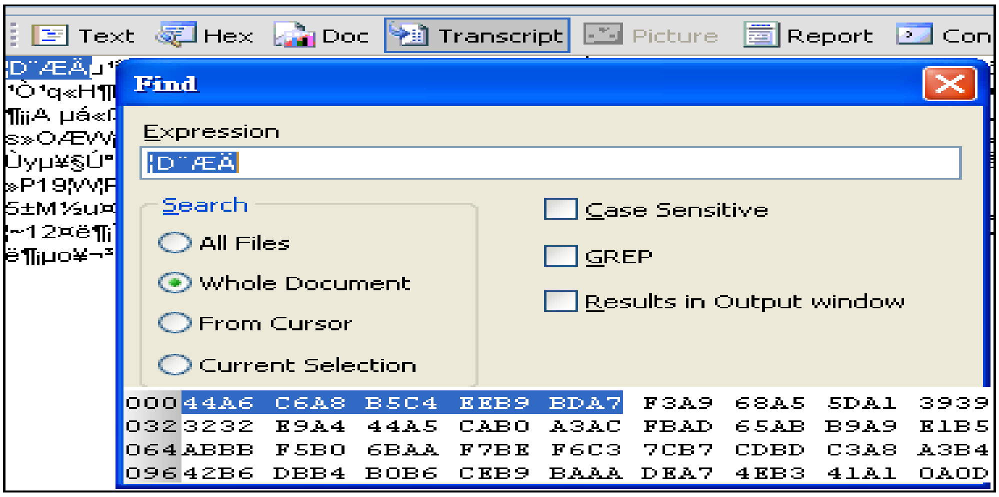
Evidence acquisition is concerned with the collection of evidence from digital devices (ex., computers, PDAs, cell phones, etc.) for extraction, examination, analysis and presentation. It plays the critical role that the digital evidence is collected in a forensically-sound manner as well as due process of law to the legal proceedings using forensic acquisition tools (ex., FTK, EnCase, E-Detective, etc.) that do not corrupt or damage the integrity of the digital evidence and ensure its admissibility in laboratory or court. A forensic acquisition tool that may be used to access suspected files or the Internet on a live or dead system without compromising the state or logical configuration of the files that they are in the reasonable suspicion. In general, computer forensic acquisition tools have one or more methods (algorithms) for verification of data-copying standard procedure that may compare the original file archive with the bit-stream image file. For example, EnCase can compute/recompute SHA-1 and MD5 hash values of selected entries and files within the open case and export them to a comma separated value (.csv) or tab delimited file format via **Search** engine [2]. EnCase Forensic produces an exact binary duplicate (bit-stream copy or forensic copy) of the original drive or storage medium, then verifies it by generating MD5 hash values for related image files and assigning CRC values to the data. These checks and balances reveal when evidence has been tampered with or altered, helping to keep all digital evidence forensically sound for use in court proceedings or internal investigations [11], and it also relates to document the chain of evidence, or chain of custody.

The extraction process enables forensic examiners to make the suspected evidence recoverable, including data viewing, keyword searching, decompressing, and carving, decrypting and bookmarking functions. A most common task in forensic examiners in extraction is searching for and recovering probative evidentiary facts. Computer forensics programs, for example, supported by EnCase EnScript, have functions for searching keywords of related criminal case during investigation. Using a **Keyword Search** will speed up the analysis process for criminal investigators or forensic examiners, if correctly; however, a poor search of keywords generates too much information or false-positive hits.

2.4 Data Viewing for Different Encoding Systems

In computer science, the terms character encoding, character map, character set or code page were historically synonymous, as the same standard would specify a repertoire of characters and how they were to be encoded into a stream of code units — usually with a single character per code unit. These terms now are related but distinct meanings, reflecting the efforts of standard bodies to use precise terminology when writing about and unifying many different encoding systems. A code page usually means a byte oriented encoding, but with emphasis on some suite of encodings (covering different scripts), where many characters share same codes in all these code pages [6].

Data viewing in different encoding systems may be the most challenging of all forensics tools. A poor data viewing, shown in Figure 3 and Figure 4 of applying EnCase **Find** function with Big-5(“刑事警察局”) code of hexadecimal values by 44A6 C6A8 B5C4 EEB9 BDA7 in a low-high byte order of RAM memory under little endian arrangement of physical address, will have a bad representation, meaninglessness, and inconsistent search/find for forensic examiners or investigators in seeking the clue in storage medium or RAM. The latest version (6.18.0.59) 2010 of the EnCase Forensic software has provided many code pages for data viewing, including ANSI Latin-1, Unicode (Little Endian), Big Endian Unicode, UTF-8, UTF-7, but does not support Big-5 and ANSI, shown in Table 2.



```
¦D¨ÆÄµ¹
¹Ò¹q«H¶
¶¡¡Â µá«
s»OÆW
Ùyµ¥§Ú
»P19
5±M½u¤
¦~12¤ë¶¡
ë¶¡µo¥¬³
```

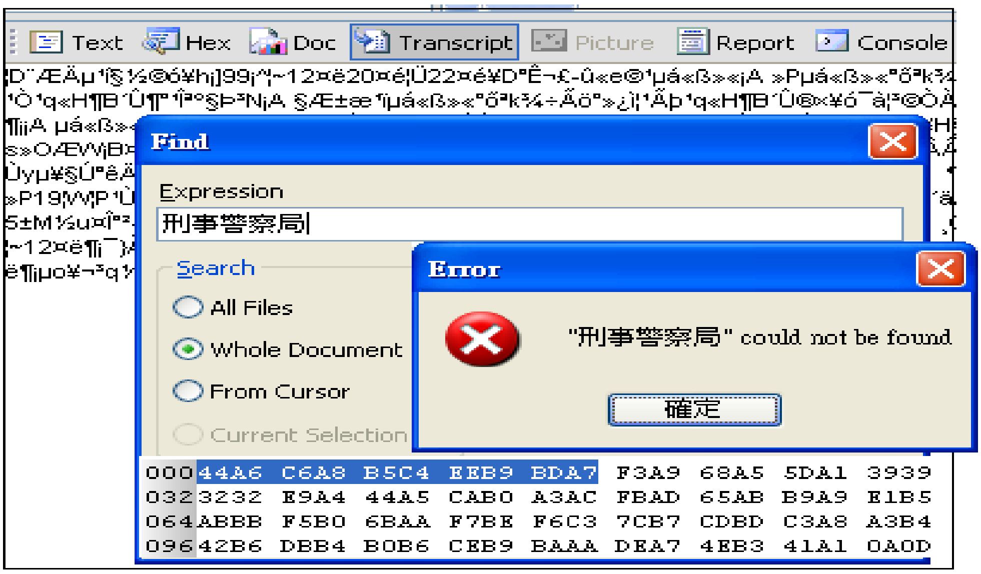
Expression: ¦D¨ÆÄ

Search
○ All Files
● Whole Document
○ From Cursor
○ Current Selection

□ Case Sensitive
□ GREP
□ Results in Output window

```
000 44A6  C6A8  B5C4  EEB9  BDA7  F3A9  68A5  5DA1  3939
032 3232  E9A4  44A5  CAB0  A3AC  FBAD  65AB  B9A9  E1B5
064 ABBB  F5B0  6BAA  F7BE  F6C3  7CB7  CDBD  C3A8  A3B4
096 42B6  DBB4  B0B6  CEB9  BAAA  DEA7  4EB3  41A1  0A0D
```

Figure 3. The “¦D¨ÆÄ” in Express field has matched in the View Pane Transcript Tab.



```
¦D¨ÆÄµ¹§½©ó¥h¡]99¡^¦~12¤ë20¤é¦Ü22¤é¥D°Ê¬£-û«e©¹µá«ß»«¡A »Pµá«ß»«°õ³k¾
¹Ò¹q«H¶B¨Û¶º¹Î³°§Þ³N¡A §Æ±æ¹µá«ß»«°õ³k¾÷Ãö°»¿¦¹Ãþ¹q«H¶B¨Û®×¥ó¯à¦³©ÒÀ
¶¡¡Â µá«ß»
s»OÆW¡B
Ùyµ¥§Ú°êÄ
»P19¡VW¡P¹Ù
5±M½u¤Î³
¦~12¤ë¶¡¯}
ë¶¡µo¥¬³q½
```

Expression: 刑事警察局

Search
○ All Files
● Whole Document
○ From Cursor
○ Current Selection

"刑事警察局" could not be found

```
000 44A6  C6A8  B5C4  EEB9  BDA7  F3A9  68A5  5DA1  3939
032 3232  E9A4  44A5  CAB0  A3AC  FBAD  65AB  B9A9  E1B5
064 ABBB  F5B0  6BAA  F7BE  F6C3  7CB7  CDBD  C3A8  A3B4
096 42B6  DBB4  B0B6  CEB9  BAAA  DEA7  4EB3  41A1  0A0D
```

Figure 4. The “刑事警察局” in Express field does not be found.

Table 2. The comparison of data viewing for different functions and encoding systems in the EnCase Forensic environment.

Encodings Functions	Notepad ANSI	Notepad Unicode	Notepad Big endian Unicode	UTF8	Other text documents Ex., .csv, .data, .arff…
View Pane Transcript Tab	Invalid	Valid	Valid	Valid	Valid, Modified*
Search	Invalid	Valid	Valid	Valid	Valid, Modified*
EnScript Program	Invalid	Valid	Valid	Valid	Valid, Modified*
File, Archive**	Invalid	Valid	Valid	Valid	Valid, Modified*
File, Delete, Archive**	Invalid	Valid	Valid	Valid	Valid, Modified*
File, Delete, Overwritten, Archive**	Invalid***	Invalid***	Invalid***	Invalid ***	Invalid***

*Insert or embed much additional information into file header in the View Pane Tabs which it is difficult understood to user.

Perform data acquisition via **Add Device function from one USB.

*** Residues of fragments may be found, or entirely overwritten by other file(s).

2.5 EnCase EnScript Language

The EnScript language supported by EnCase software is a scripting language and application program interface (API). EnScript and customized program design will allow forensic examiners to quickly search and retrieve any relevant data and provide for further analysis with built-in EnScript or by developing their own EnScript tools [4]. In any EnCase application, EnScript programs are organized and stored in the EnScript library. EnScript programs allow investigators and programmers to develop utilities to automate and facilitate forensic investigations. The source code of programs stored in either 8-bit text or Unicode can be interpreted and shared with other investigators.

A scripting language EnScript, such as VBA for Microsoft Excel or Microsoft Word of macro-programming language, an object-oriented programming (OOP) language with combination of Java and C++ operational mechanism as well as characteristics, is designed to allow programmers with some programming techniques (especially OOP) to fully develop useful functions related to the range of the digital evidence in the EnCase Forensic environment, to automate time-consuming investigative tasks, such as searching and analyzing specific document types or other labor-intensive processes and procedures, and to create functionally customizable applications. EnScript has evolved into a powerful object oriented language with case, inheritance, virtual functions, type reflection and a threading model. EnScript also supports COM libraries from other applications, allowing programmers to automate document processing tasks and remote data retrieval through DCOM [3]. For example, programmers can directly call Microsoft Office VBA macro programs with or without parameters and running them via **app.Run**("**VBAModuleName.SubOrFunctionName**", **parameters…)** syntax format and COM libraries.

3. Methodology

3.1 Software and Hardware Requirements

In this study, the software and hardware requirements are as follows:

(1) Personal Computer with x86 platform;

(2) Universal Serial Bus (USB) device with about 2.0GB storage volume;

(3) Microsoft Office Word and Excel 2010 version;

(4) Source programs for digital evidence extraction: ES_InternalCode.EnScript, ES_MIS2011_AddDevice.EnScript (See Appendix), EnCaseMacroCallEnd (VBA), EnCaseMacroCallSave (VBA), and EnCaseMacroCallQuit (VBA), and EnCaseWordFileClose (VBA);

(5) Notepad;

(6) Guidance Software (U.S.) EnCase® Forensic 2010 6.18.0.59 version with one registered dongle device.

3.2 Configuration of Digital Evidence Automation Extraction

Figure 5 shows the profile configuration of digital evidence extraction. It includes nine main steps:

(1) Create four text files and save them with ANSI, Unicode (little endian), Unicode big endian, and UTF-8 encode, respectively.

(2) Create one Excel file and one Word file.

(3) Create or copy files with file types of .txt, .data, .arff, csv, and .mht.

(4) Create one case file of the EnCase Forensic to store these files into **Single Files**.

(5) Create an exact image (bit-stream copy) of one USB disk with 2.0 GB of live acquisition from **Add Device** function of the EnCase Forensic.

(6) Execute EnScript and VBA programs.

(7) Convert these files with different encode into meaningful contents via **File Conversion** of Microsoft Office Word.

(8) Store contents of processing file in **MemoryFileClass** object and then proceed to search keywords via **SearchClass**.

(9) Finally, save the results to one file of Word document.

A type library (**typelib** Excel/Word shown in Figure 5) is a binary file containing all the type information programmers need to use procedures or classes in DLLs [6]. Type information contains a description of everything a client needs to know to use an object’s service. For example, the type information of an object contains a list of the interface’s methods (also called functions) and properties, along with a description of the parameters for those methods (functions) [7].

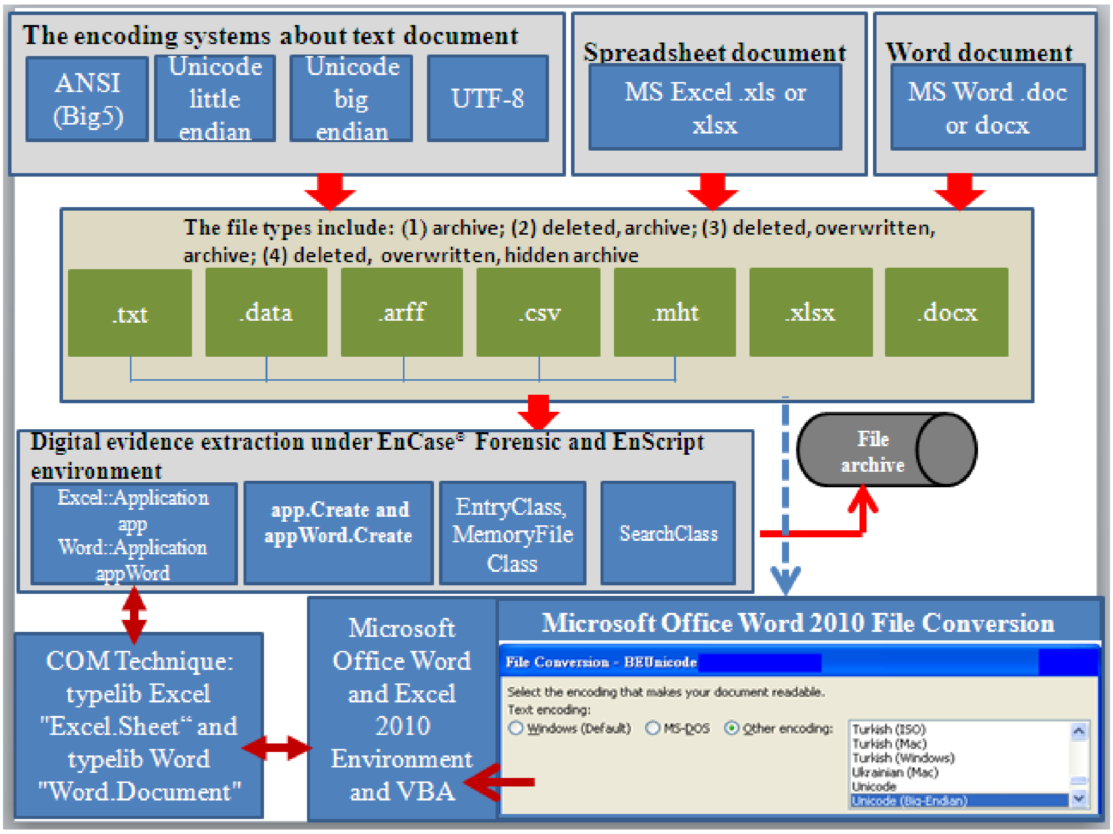

The encoding systems about text document
ANSI (Big5)
Unicode little endian
Unicode big endian
UTF-8
Spreadsheet document
MS Excel .xls or xlsx
Word document
MS Word .doc or docx
The file types include: (1) archive; (2) deleted, archive; (3) deleted, overwritten, archive; (4) deleted, overwritten, hidden archive
.txt
.data
.arff
.csv
.mht
.xlsx
.docx
Digital evidence extraction under EnCase® Forensic and EnScript environment
Excel::Application app Word::Application appWord
app.Create and appWord.Create
EntryClass, MemoryFile Class
SearchClass
File archive
COM Technique: typelib Excel "Excel.Sheet“ and typelib Word "Word.Document"
Microsoft Office Word and Excel 2010 Environment and VBA
Microsoft Office Word 2010 File Conversion
File Conversion - BEUnicode
Select the encoding that makes your document readable.
Text encoding:
○ Windows (Default) ○ MS-DOS ● Other encoding:
Turkish (ISO)
Turkish (Mac)
Turkish (Windows)
Ukrainian (Mac)
Unicode
Unicode (Big-Endian)

Figure 5. The profile configuration of digital evidence extraction via COM technique.

3.3 Experimental Samples

Table 3 is shown the experimental samples, their contents referred to http://www.cib.gov.tw/index.aspx （內政部警政署刑事警察局全球資訊網), which provided for the purpose of digital evidence extraction via **Add Device** live acquisition and **Single Files**, and its operational environment of EnCase Forensic is shown in Figure 6 and Figure 7, respectively.

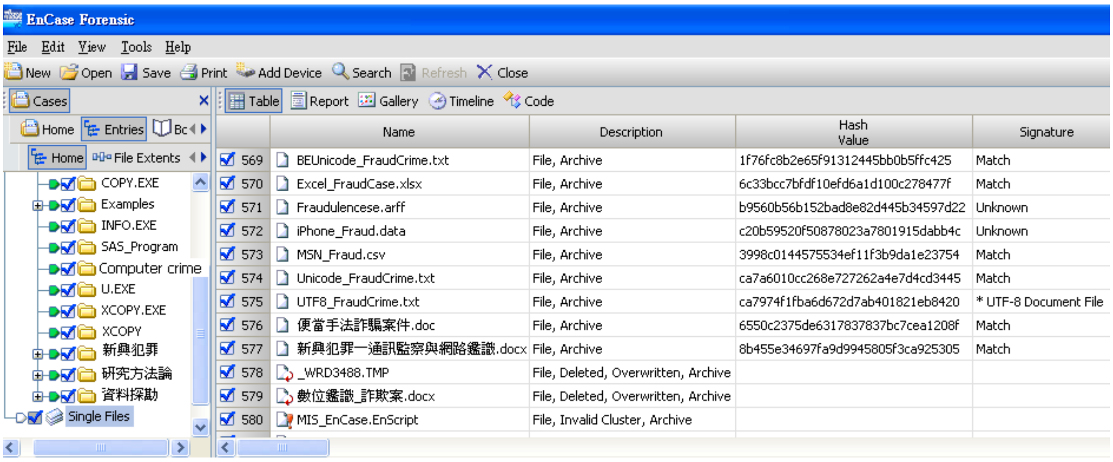
Table 3. The experimental samples for digital evidence extraction.

File Name	File Type	Encode System	File Description	Hash Value (MD5)
_WRL3836.TMP	Windows Temporary	ANSI(Big5) and Unicode	File, Deleted, Overwritten, Hidden Archive	N/A *
Computer Crime.rtf	Rich Text Format (.rtf)	ANSI(Big5)	File, Deleted, Overwritten, Archive	N/A*

Table 3. The experimental samples for digital evidence extraction (continued).

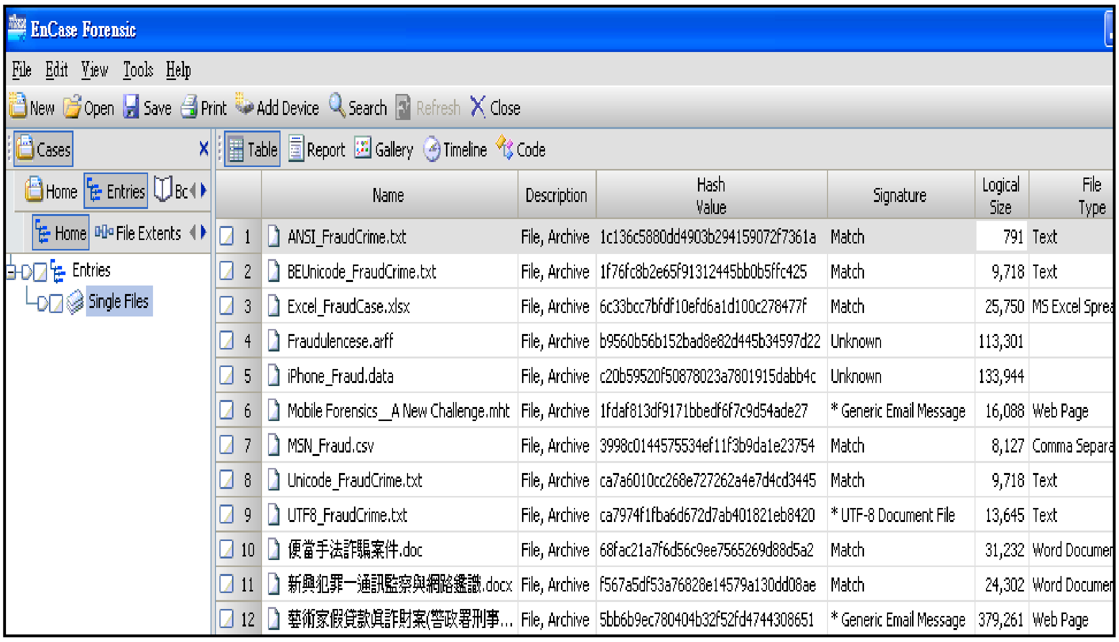
File Name	File Type	Encode System	File Description	Hash Value (MD5)
_ô¤fºÊ~1.DOC	Word Document (.doc)	ANSI (Big5) and Unicode little Endian	File, Deleted, Archive	031040151b4ef9d0a6dbb8c924ec6039
ANSI_FraudCrime.txt	Text (.txt)	ANSI	File, Archive	1c136c5880dd4903b294159072f7361a
BEUnicode_FraudCrime.txt	Text (.txt)	Unicode Big Endian	File, Archive	1f76fc8b2e65f91312445bb0b5ffc425
Excel_FraudCase.xlsx	MS Excel Spreadsheet(.xlsx)	ANSI (Big5) and Unicode little Endian	File, Archive	6c33bcc7bfdf10efd6a1d100c278477f
Fraudulencese.arff	Unknown(.arff)	ANSI	File, Archive	b9560b56b152bad8e82d445b34597d22
iPhone_Fraud.data	Unknown(.data)	ANSI	File, Archive	c20b59520f50878023a7801915dabb4c
MSN_Fraud.csv	Comma Separated Database	ANSI	File, Archive	3998c0144575534ef11f3b9da1e23754
Unicode_FraudCrime.txt	Text (.txt)	Unicode Little Endian	File, Archive	ca7a6010cc268e727262a4e7d4cd3445
UTF8_FraudCrime.txt	Text (.txt)	UTF-8	File, Archive	ca7974f1fba6d672d7ab401821eb8420
便當手法詐騙案件.doc	Word Document (.doc)	ANSI (Big5) and Unicode little Endian	File, Archive	68fac21a7f6d56c9ee7565269d88d5a2
新興犯罪－通訊監察與網路鑑識.docx	Word Document (.doc)	ANSI (Big5) and Unicode little Endian	File, Archive	f567a5df53a76828e14579a130dd08ae
藝術家假貸款真詐財案(警政署刑事警察局).mht	Web Page (.mht)	Unicode little Endian and UTF-8	File, Archive	5bb6b9ec780404b32f52fd4744308651
Mobile Forensics__A New Challenge.mht	Web Page (.mht)	Unicode little Endian and UTF-8	File, Archive	1fdaf813df9171bbedf6f7c9d54ade27

* It cannot produce one hash value of MD5 under a deleted and overwritten (hidden) file.



	Name	Description	Hash Value	Signature
☑ 569	BEUnicode_FraudCrime.txt	File, Archive	1f76fc8b2e65f91312445bb0b5ffc425	Match
☑ 570	Excel_FraudCase.xlsx	File, Archive	6c33bcc7bfdf10efd6a1d100c278477f	Match
☑ 571	Fraudulencese.arff	File, Archive	b9560b56b152bad8e82d445b34597d22	Unknown
☑ 572	iPhone_Fraud.data	File, Archive	c20b59520f50878023a7801915dabb4c	Unknown
☑ 573	MSN_Fraud.csv	File, Archive	3998c0144575534ef11f3b9da1e23754	Match
☑ 574	Unicode_FraudCrime.txt	File, Archive	ca7a6010cc268e727262a4e7d4cd3445	Match
☑ 575	UTF8_FraudCrime.txt	File, Archive	ca7974f1fba6d672d7ab401821eb8420	* UTF-8 Document File
☑ 576	便當手法詐騙案件.doc	File, Archive	6550c2375de6317837837bc7cea1208f	Match
☑ 577	新興犯罪一通訊監察與網路鑑識.docx	File, Archive	8b455e34697fa9d9945805f3ca925305	Match
☑ 578	_WRD3488.TMP	File, Deleted, Overwritten, Archive		
☑ 579	數位鑑識_詐欺案.docx	File, Deleted, Overwritten, Archive		
☑ 580	MIS_EnCase.EnScript	File, Invalid Cluster, Archive		

Figure 6. The operational environment of the experimental samples via Add Device live acquisition.


EnCase Forensic
File Edit View Tools Help
New Open Save Print Add Device Search Refresh Close
Cases
Table Report Gallery Timeline Code
Home Entries Bc
Home File Extents
Entries
Single Files

	Name	Description	Hash Value	Signature	Logical Size	File Type
□ 1	ANSI_FraudCrime.txt	File, Archive	1c136c5880dd4903b294159072f7361a	Match	791	Text
□ 2	BEUnicode_FraudCrime.txt	File, Archive	1f76fc8b2e65f91312445bb0b5ffc425	Match	9,718	Text
□ 3	Excel_FraudCase.xlsx	File, Archive	6c33bcc7bfdf10efd6a1d100c278477f	Match	25,750	MS Excel Sprea
□ 4	Fraudulencese.arff	File, Archive	b9560b56b152bad8e82d445b34597d22	Unknown	113,301	
□ 5	iPhone_Fraud.data	File, Archive	c20b59520f50878023a7801915dabb4c	Unknown	133,944	
□ 6	Mobile Forensics__A New Challenge.mht	File, Archive	1fdaf813df9171bbedf6f7c9d54ade27	* Generic Email Message	16,088	Web Page
□ 7	MSN_Fraud.csv	File, Archive	3998c0144575534ef11f3b9da1e23754	Match	8,127	Comma Separa
□ 8	Unicode_FraudCrime.txt	File, Archive	ca7a6010cc268e727262a4e7d4cd3445	Match	9,718	Text
□ 9	UTF8_FraudCrime.txt	File, Archive	ca7974f1fba6d672d7ab401821eb8420	* UTF-8 Document File	13,645	Text
□ 10	便當手法詐騙案件.doc	File, Archive	68fac21a7f6d56c9ee7565269d88d5a2	Match	31,232	Word Documen
□ 11	新興犯罪一通訊監察與網路鑑識.docx	File, Archive	f567a5df53a76828e14579a130dd08ae	Match	24,302	Word Documen
□ 12	藝術家假貸款真詐財案(警政署刑事...	File, Archive	5bb6b9ec780404b32f52fd4744308651	* Generic Email Message	379,261	Web Page

Figue 7. The operational environment of the experimental samples via Single Files.

4. Results and Discussion

4.1 Results

In this paper, it will apply the most popular fraud cases in Taiwan to construct keywords, including "刑事", "檢察官", "詐欺", "菲律賓" and "Mobile", and results of searching hits for these experimental files are shown in Table 4. And we have gotten remarkable success of results in completely searching hits as reviewing contents of experimental samples.

Table 4. The results of running ES_InternalCode.EnScript, ES_MIS2011_ AddDevice.EnScript, and VBA programs.

File Name	Search Hits of Keywords				
	刑事	檢察官	詐欺	菲律賓	Mobile
_WRL3836.TMP*	Meaningless	Meaningless	Meaningless	Meaningless	Meaningless
Computer Crime.rtf*	Meaningless	Meaningless	Meaningless	Meaningless	Meaningless
_ô¤fºÊ~1.DOC*	3	1	2	3	0
ANSI_FraudCrime.txt**	2	1	7	7	0
BEUnicode_FraudCrime.txt **	28	3	46	85	0
Excel_FraudCase.xlsx**	3	1	2	1	0
Fraudulencese.arff**	2	3	1	0	0
iPhone_Fraud.data**	3	1	0	0	0
MSN_Fraud.csv**	2	0	3	0	0
Unicode_FraudCrime.txt**	29	**4**	43	79	1
UTF8_FraudCrime.txt**	17	**4**	35	53	4
便當手法詐騙案件.doc**	1	0	1	0	1
新興犯罪－通訊監察與網路鑑識.docx**	3	1	5	0	0
藝術家假貸款真詐財案(警政署刑事警察局).mht**	3	1	1	0	0
Mobile Forensics__A New Challenge.mht**	0	0	0	0	4

*Files obtained from one USB via operation of **Add Device** function. They must be stored in hard disk of local machine before proceeding keywords searching.

Files obtained from hard disk of local machine via operation of **Single Files function.

4.2 Discussion

In the beginning, we employ these classes of **EntryClass** and **EntryFileClass** to get files from **Single Files**, and **CodePageClass** to set encode of files, to obtain the correct encoding from these original files, but some problems are arisen, such as:

(1) From ANSI (Big-5 code) of Notepad file we cannot get the correct content.

(2) By using **EntryFileClass.ReadString (text, -1, "\x0d\x0a")** to get one paragraph contents every time, some contents of this paragraph may be lost occasionally to full-pitch characters.

(3) The use of **text.Contains ("Keyword")** search in a piecewise way on a large number of files in high storage space will result in significantly increasing the execution time.

(4) In the file types of text document, for example, .data, .csv, .arff , or others, there is some additional information that is embedded into the **Single Files** and data viewing, and may also interfere with the accuracy of keywords searching.

(5) **EntryClass** and **EntryFileClass** can only process MS file systems in storage media, such as FAT or NTFS, and image file from bit-steaming live acquisition cannot be managed and allocated by FAT or NTFS.

(6) Excel file cannot be properly processed by keyword search.

To solve the above problems, some amendment methods are proposed in the following:

(1) Microsoft Office Word provides a function of file conversion to convert different encoding files into properly internal code and correctly display content in Word's editing area. We should prepare file ready in advance before keyword searching is processed. By this way we can solve the incompatibility and conflict about different encoding and types of file.

(2) As for Microsoft Office Excel file, we may open it by VBA macro program, copy the entire contents to the clipboard, and then paste it down on edit area of Word.

(3) We can get all contents from Word via COM and store them to one string buffer, open this buffer by **MemoryFileClass**, and finally speed up keyword search by **SearchClass** and calculate search hits.

(4) The frequent actions of opening, processing and closing files may directly affect the execution performance. In order to solve this problem, it is necessary to set **SetVisible()** to **false** during loading and processing files of Word or Excel.

(5) We must save any image generated from live acquisition into files in the physical drive (such as FAT or NTFS) so that we can execute programs on these files. We may use **LocalFileClass**, **EntryFileClass**, **DocumentClass** and **TranscriptFileClass** for that purpose.

4.3 Shortcomings of EnScript Language

Although the EnScript language of EnCase Forensic software has provided the very powerful functions for acquisition and extraction of digital evidence, there are still some shortcomings, including that:

(1) It provides only few functions for statistical analysis, so we must use more statistical functions in the Excel via COM technique.

(2) So far, it still cannot provide the related classes for the Excel files in the case file as it is necessary to fetch and process them.

(3) The correct contents cannot be obtained from Excel files stored in the case file via **DocumentClass:: TranscriptFileClass**, and may be lost when data type is a numeric value.

(4) Until now, it does still not provide syntax of type library for VBA Word and Excel in EnCase for users through browsing **Help** menu of EnCase. Therefore, program designers must be forced to try and error by chance while running programs; however, there is some difference in syntax structures of VBA between EnCase and Microsoft Word/Excel.

(5) It provides two call methods, namely, call by value and call by name. Although the calling program of EnScript calls the called one of VBA for Word or Excel with mechanisms of call by value and call by reference via COM, it doesn't allow to change any parameter's value via call by reference. The problem may be solved by using a function call and return only one value. If it is necessary to return multiple values, you must use **selection.TypeText()** and **selection. MoveRight(unit,count,extend)** macro statements in the Word environment. In Excel environment, you must temporally store at one or more cells, and then use **sheet. Range("range").Value()** or **cells.Item (row, col)** macro statements.

(6) The macro functions of VBA for Word and Excel may not fully be supported by the EnScript language.

(7) In the EnScript environment, we must compile and debug the calling/called programs, respectively. The main reason is that EnScript cannot accept the called programs of Word/Excel VBA if there is any mistake or bug, and no results or error message is present if there is error.

5. Conclusion

In this paper we have presented a practical method by COM technique to solve the bad representation of data viewing and keyword searching in extraction phase, which is caused by the conflict of different encode and types of files, and to get good results to search hits. EnCase software is an integrity forensics tool with very powerful functions. It has obtained a legitimate tool approved by court and law enforcement agency in United States and other countries around the world. For the past few years EnCase functions (including EnScript language) have been improved incessantly and its updated version is actively downloaded via website of Guidance Software Inc. Therefore it will enable more law enforcement officers and private enterprise information security employees to use.

The material related to EnScript language and VBA we can refer to is very limited. In the future, we expect there will be further understanding and practical application, especially, on how to use COM technique for automatic tasks of forensic standard operating procedures (SOPs) on digital evidence.

In the future, we will focus on the below topic:

(1) Developing the practical automation software system align to the standard

operating procedures (SOPs) on digital evidence collection of law enforcement agency, for example, the production of criminal case documents.

(2) Developing the techniques of sniff (or intercept), mirror, and packets interception over wired and wireless network on Internet (ex., E-Detective production), and then feeding collected data into case files of EnCase and linking the COM technology with them.

(3) Integrating EnCase, Excel with crime link analysis software (ex., i2 Analyst's laptop appliance), so that we can build social network among suspects and reconstruct modus operandi related to cyber crime.

Acknowledgements

The work of this paper has supported in part by Decision Group Inc., Taipei, Taiwan, 2009/6/1-2010/5/31.

References

[1] Bill Nelson, Amelia Phillips, and Christopher Stuart, *Guide to Computer Forensics and Investigations*, Course Technology, Cengage Learning, 2010, pp.261-272.

[2] EnCase Forensic Version 6.14.1 User’s Guide, Guidance Software Inc., 2009.

[3] EnCase V5 EnScript Language Reference , Guidance Software Inc., 2005.

[4] Eoghan Casey, *Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet*, Elsevier Academic Press, 2nd Edition, 2004, p.668.

[5] http://cc.ee.ntu.edu.tw/~skjeng/, 2010/11/12.

[6] http://en.wikipedia.org/wiki/Character_encoding, 2011/3/11.

[7] http://en.wikipedia.org/wiki/Digital_evidence, 2010/12/30.

[8] http://en.wikipedia.org/wiki/Locard's_exchange_principle, 2011/3/4.

[9] http://msdn.microsoft.com/en-us/library/ms809980.aspx, 2010/12/30.

[10] http://vb.mvps.org/hardcore/html/whatistypelibrary.htm, 2011/2/3.

[11] http://www.guidancesoftware.com/forensic.htm, 2011/3/10.

Appendix

```
/* File name : ES_MIS2011_AddDevice.EnScript
   Department: Department of Information Management, Central Police University (CPU), Taiwan
   Author : Professor Wu, Kuo-Ching    */
typelib Word "Word.Document"
typelib Excel "Excel.Sheet"
#ifdef Word
class MainClass {
  String 路徑名稱 ;
  int nCount;
  void Main(CaseClass c) {
    路徑名稱 = "E:\\EnCaseTEMP";//預設, 路徑名稱必須存在於儲存設備內
    nCount = 0;
    SystemClass::ClearConsole(); //clear console contents
    Console.WriteLine ("\t\t\t 這個 Case 案件檔案 " + c.Name() + " 含有下列數位證據檔案");
    Console.WriteLine ("=================================================");
    對話盒路徑類別 對話盒輸入路徑(null, this);
    if (SystemClass::OK == 對話盒輸入路徑.Execute()) {
      forall (EntryClass e in c.EntryRoot()) {
        if ( e.Name() == "Single Files") continue;
        if (e.IsSelected() && !e.IsFolder() && !e.IsUnallocated() && !e.IsSparse())
        {
          nCount++;
          Console.WriteLine ("\t\t*** [ " + nCount + ". " + e.Name() + " ] filename. ***");
          AddDeviceFilesSaveClass clsProcessFile;
          clsProcessFile.SaveProcessFile(e, 路徑名稱, nCount);
          Console.WriteLine ("\n");
        }//end if IsSelected
      }//forall
      if (nCount == 0)    {
        SystemClass::Message(0, "Not Selected V file(s)",
            "[ERROR]: Please V Select file(s) from this Case obtain from [Add
Device](USB)function!");
      }//end if nCount
      Console.WriteLine ("\n\n ====== Program has finished. =========");
    }//if SystemClass
  }//Main
 }//MainClass
 class AddDeviceFilesSaveClass {
```

```
//functions
void SaveProcessFile(EntryClass e, const String&  路徑名稱, const int& nCount) {
   LocalFileClass OutputTranscriptFile();
   Console.WriteLine("Selected file name: " + e.FullPath() + ", and Logical Size: " +
  e.LogicalSize() + " Byte(s).");
   String outFilename, strMemoryBuffer;
   uint count 刑事     = 0, count 檢察官  = 0, count 詐欺     = 0, count 菲律賓  = 0,
          countMobile = 0;
   outFilename.BuildPath(路徑名稱, "EnCase$" + e.Name());
   EntryFileClass theEntryFile();
   theEntryFile.Open(e);
   DocumentClass WordDoc();
   DocumentClass::ProcessOptions docProc = WordDoc.ProcessMethod(e);
   Console.WriteLine("\t\tYou IsSelected Entry Name: " +
   DocumentClass::ProcessOptions::SourceText(docProc));
   DocumentClass::TranscriptFileClass transcriptFile();
   OutputTranscriptFile.Open(outFilename, FileClass::WRITE);
   OutputTranscriptFile.WriteBuffer(transcriptFile.File, -1);
   OutputTranscriptFile.Close();
   Console.WriteLine("\t\t\t File will save: " + outFilename + ", and Logical Size: " +
  transcriptFile.File.GetSize() + " Byte(s).");
             Word::Application appWord;
             if (appWord.Create()) {
                 appWord.SetVisible(false);
                 Word::Documents docs = appWord.Documents();
                 Word::Document doc = docs.Add();
                 Word::Range range = doc.Range();
                 Word::Selection selection = appWord.Selection();
                 appWord.Run("載入 EnCase 檔案", outFilename);
                 strMemoryBuffer.Close ();
                 strMemoryBuffer = doc.Range().Text();//assign the entire document to str
                 Console.WriteLine (strMemoryBuffer);
                 appWord.Run("EnCaseWordFileClose");
             }//if-appWord.Create
             else {
                 SystemClass::Message(0, "MS Word or Excel Not Available",
                 "[ERROR]: Please Plug-in EnCase Dongle before starting EnCase software!");
                 SystemClass::Exit();
```

```
      }//end    WordApp.Create
   MemoryFileClass file();
   if (file.Open(strMemoryBuffer)) {
     SearchClass search();
     search.AddKeyword("刑事",      KeywordClass::UNICODE);
     search.AddKeyword("檢察官", KeywordClass::UNICODE);
     search.AddKeyword("詐欺",      KeywordClass::UNICODE);
     search.AddKeyword("菲律賓", KeywordClass::UNICODE);
     search.AddKeyword("Mobile", KeywordClass::UNICODE);
     count 刑事      = 0, count 檢察官  = 0, count 詐欺      = 0, count 菲律賓  = 0;
     countMobile = 0;
     if (search.Create()) { // Initialize the search engine
         uint count = search.Find(file); // Search the file
         if (count > 0) {
             SearchClass::HitArrayClass hits = search.GetHits(); //
             foreach (SearchClass::HitClass hit in hits) { // Itera
                 if (hit.KeywordIndex() == 0)         // KeywordIndex() i
                     ++count 刑事;
                 else if (hit.KeywordIndex() == 1)
                     ++count 檢察官;
                 else if (hit.KeywordIndex() == 2)
                     ++count 詐欺;
                 else if (hit.KeywordIndex() == 3)
                     ++count 菲律賓;
                 else if (hit.KeywordIndex() == 4)
                     ++countMobile;
              }//foreach-hit
          }//if-count
     }//if-search.Create
   }//if-MemoryFileClass
   Console.WriteLine ("\t\t\t\t\tcount 刑事      = " + count 刑事);
   Console.WriteLine ("\t\t\t\t\tcount 檢察官  = " + count 檢察官);
   Console.WriteLine ("\t\t\t\t\tcount 詐欺      = " + count 詐欺);
   Console.WriteLine ("\t\t\t\t\tcount 菲律賓  = " + count 菲律賓);
   Console.WriteLine ("\t\t\t\t\tcountMobile = " + countMobile);
   Word::Application WordApp;
   if (WordApp.Create()) {
        WordApp.SetVisible(false);
```

```
                Word::Documents docs = WordApp.Documents();
                Word::Document docX = docs.Open("K:\\數位鑑識_詐欺案.docx");
                Word::Selection selection = WordApp.Selection();
                WordApp.Run("EnCaseMacroCallEnd");
                selection.TypeText( nCount + ".  檔案名稱  = " + e.OriginalPath());
                selection.TypeText("\n");
                selection.TypeText("\t\t 刑事      Hit  次數      = " + count 刑事);
                selection.TypeText("\n");
                selection.TypeText("\t\t 檢察官  Hit  次數      = " + count 檢察官);
                selection.TypeText("\n");
                selection.TypeText("\t\t 詐欺      Hit  次數      = " + count 詐欺);
                selection.TypeText("\n");
                selection.TypeText("\t\t 菲律賓  Hit  次數      = " + count 菲律賓);
                selection.TypeText("\n");
                selection.TypeText("\t\tMobile Hit  次數      = " + countMobile);
                selection.TypeText("\n");
                WordApp.Run("EnCaseMacroCallSave");//call macro of VBA by COM
                WordApp.Run("EnCaseMacroCallQuit");
            }//if-WordApp.Create
  }//SaveProcessFile function
}//end class
class  對話盒路徑類別: DialogClass {
   PathEditClass PathEdit;
   對話盒路徑類別(DialogClass parent, MainClass main):
   DialogClass(parent, "Document::TranscriptClass"),
   PathEdit(this,  "輸入路徑名稱(存在)",  DialogClass::START,  DialogClass::START,  400,
DialogClass::DEFAULT, 0, main.路徑名稱, WindowClass::FOLDEROPEN) {}
}
#else
class MainClass {
  void Main(CaseClass c) {
     SystemClass::Message(0, "MS Word Not Available",
        "[ERROR]: Please Plug-in EnCase Dongle before starting EnCase software!");
  }
}
#endif
```