Advanced Evidence Collection and Analysis of Web Browser Activity

By

Junghoon Oh, Seungbong Lee and Sangjin Lee

Presented At

The Digital Forensic Research Conference

DFRWS 2011 USA New Orleans, LA (Aug 1st - 3rd)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http:/dfrws.org

Advanced Evidence Collection and Analysis of Web Browser Activity

J. Oh, S. Lee and S. Lee

Digital Forensics Research Center, Korea University

Junghoon Oh

blue0226@korea.ac.kr

Agenda

1. Introduction
2. Problem of existing research and tools
3. Advanced evidence analysis
4. Tool development
5. Conclusions and Future Research

Introduction

- ✓ **Need for web browser forensics**
  - **Most people use web browser for Using the Internet**
  - **Also, suspect use web browser for his purpose**
    - **Collect information for advance preparation**
    - **Hide his/her crime**
    - **Search for crime method**

  - **So, web browser forensics helps investigator to analyze crime act an d crime plan of suspect.**

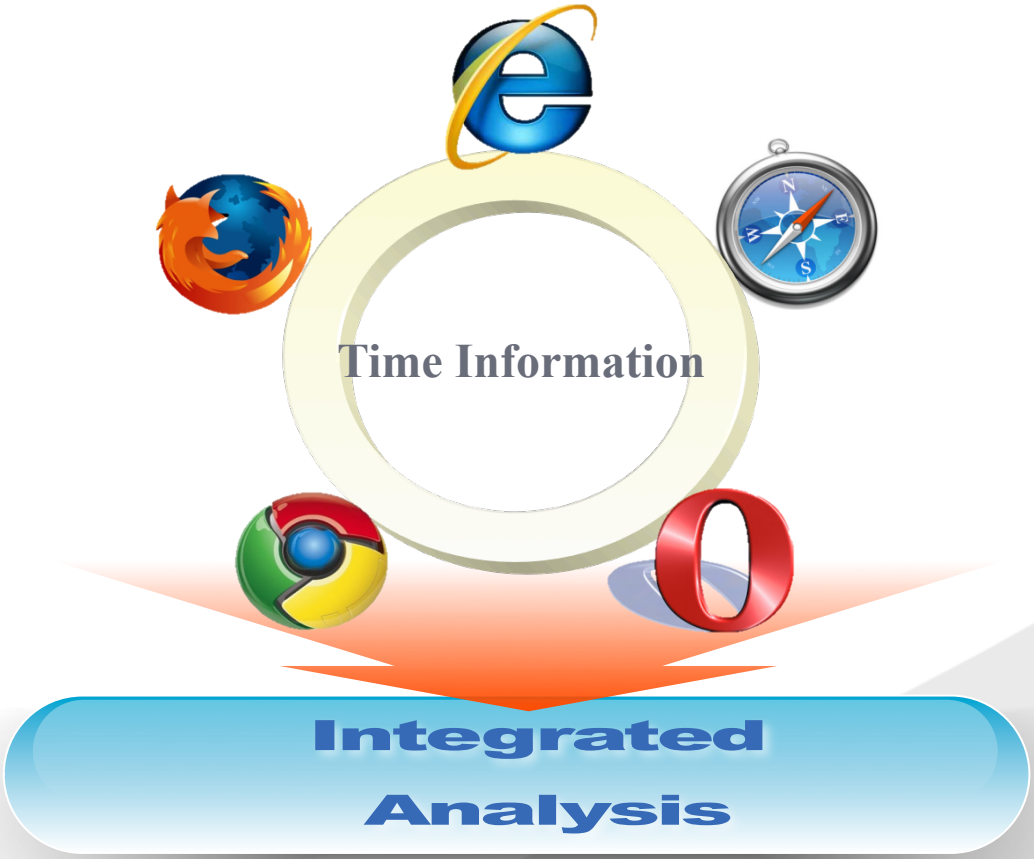
Problem of Existing Research and Tools

- ✓ **Targeting a specific web browser or a specific log file**
  - **The environment of using multiple web browser**
  - **One user can user multiple web browser in single system.**
  - **Specific web browser or log file analysis is inappropriate in web browser forensics.**

- ✓ **Simple parsing level**
  - **Simply parse the information included in log file**
  - **No extracting more significant information**
    - **Search word**
    - **User Activity**

Advanced Evidence Analysis

- ✓ **Integrated analysis**
  - **Need for integrated analysis of multiple web browser**
  - **Integrated analysis through time information**



Advanced Evidence Analysis

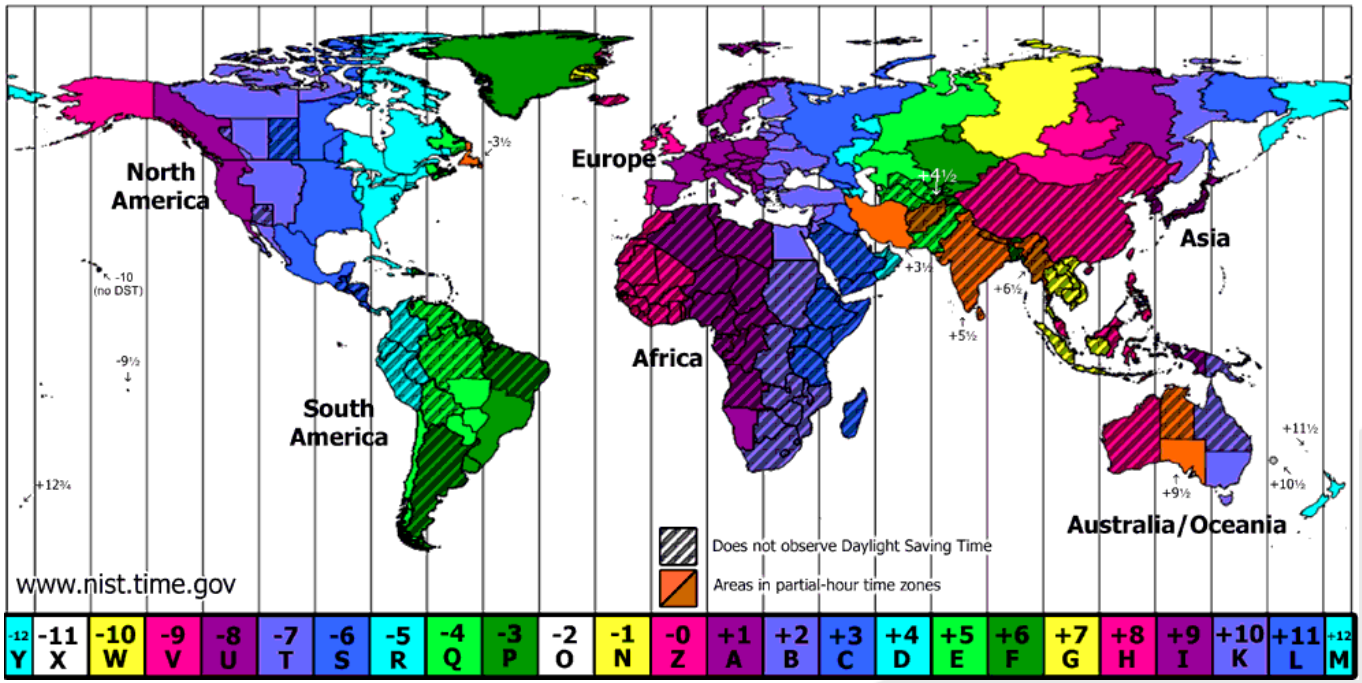
✓ Integrated analysis

– Time format used by five web browser

Web Browser	Time Format
Internet Explorer	FILETIME: 100-nanosecond (10^{-9}) Since January 1, 1601 00:00:00 (UTC)
Firefox	PRTime: microsecond(10^{-6}) Since January 1, 1970 00:00:00 (UTC)
Chrome	WEBKIT Time: microsecond(10^{-6}) Since January 1, 1601 00:00:00 (UTC)
Safari	CFAbsoluteTime: second Since January 1, 2001 00:00:00 (UTC)
Opera	UNIX Time: second Since January 1, 1970 00:00:00 (UTC)

Advanced Evidence Analysis

- ✓ **Time zone analysis**
  - **Time zone information**
    - **Five web browser's time format use UTC time**
    - **Need for convert from UTC time to Local time**



Advanced Evidence Analysis

✓ Analysis of search history

- Search word used in search engine is saved in URL

- General URL information structure

http://	Host	Port	/	Path	?	Searchpart(Variable=Value)

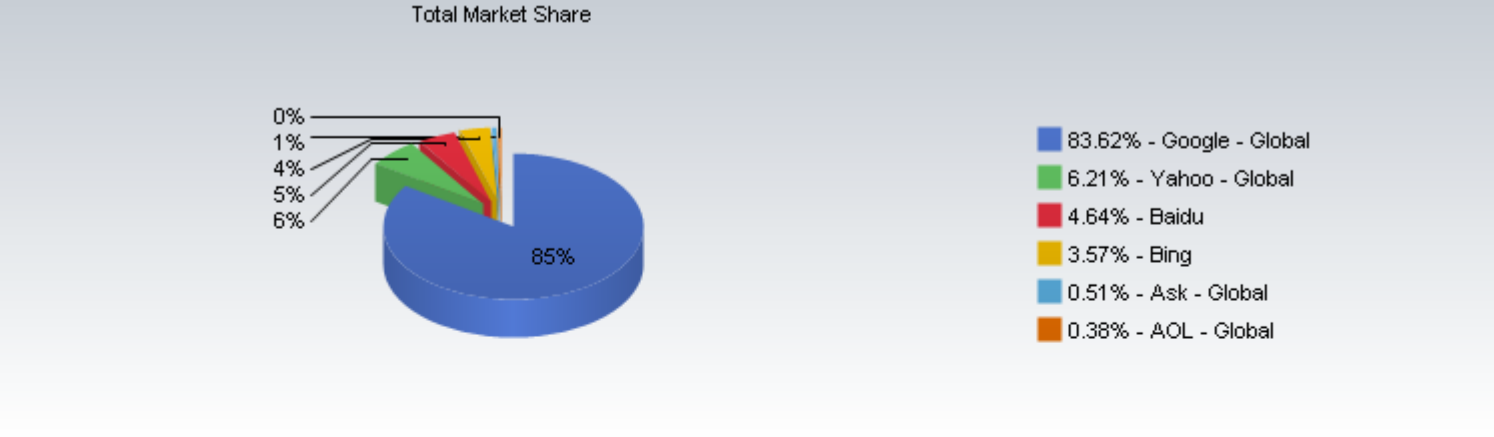
- The URL information generated when search “forensic” keyword in Google search engine

http://www.google.com/search?hl=en&source=hp&q=forensic&aq=f&oq=&aqi=g10

Advanced Evidence Analysis

✓ Analysis of search history

– **Global market share of search engines (released by NetMarketShare)**



Search Engine	Total Market Share
Google - Global	83.62%
Yahoo - Global	6.21%
Baidu	4.64%
Bing	3.57%
Ask - Global	0.51%
AOL - Global	0.38%
Excite - Global	0.03%
Lycos - Global	0.02%
Microsoft Live Search	0.00%
AltaVista - Global	0.00%
MSN - Global	0.00%

Advanced Evidence Analysis

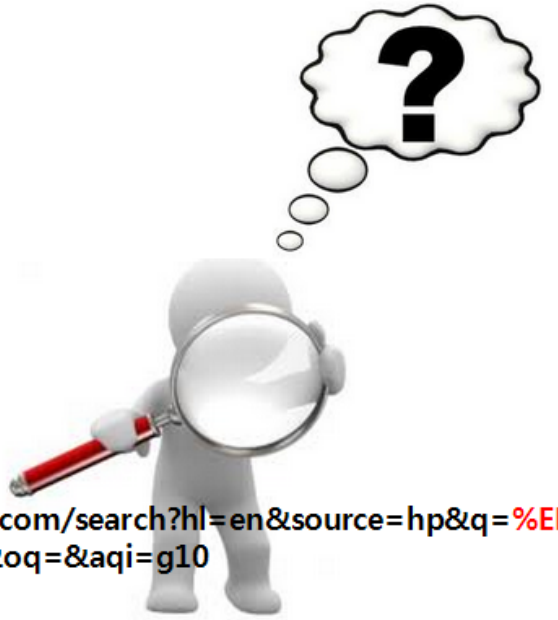
✓ Analysis of search history

- Host, path and search word location for different search engines

Search Engine	Host	Path	Search Word Location
Google	google.com	#sclient	After variable q
Yahoo	search.yahoo.com	/search	After variable p
Baidu	baidu.com	/s	After variable wd
Bing	bing.com	/search	After variable q
Ask	ask.com	/web	After variable q
AOL	search.aol.com	/search/	After variable q
Excite	msxml.excite.com	/results/	After path /Web/
Lycos	Search.lycos.com		After variable query
Alta vista	altavista.com	/search	After variable p
MSN	bing.com	/search	After variable q

Advanced Evidence Analysis

- ✓ **Analysis on URL encoding**
  - **Non-alphabetic characters are encoded in URL.**
  - **URL encoding confuse a digital forensic investigator.**



http://www.google.com/search?hl=en&source=hp&q=%ED%8F%AC%EB%A0%8C%EC%8B%9D&aq=f&oq=&aqi=g10

Advanced Evidence Analysis

✓ Analysis on URL encoding

- URL encoding
  - Hexadecimal code and Character '%'
  - Example of URL encoding

    ➔ %ED%8F%AC%EB%A0%8C%EC%8B%9D

- URL encoding method
  - Unicode Encoding
  - UTF-8 Encoding
  - Code Page Encoding

Advanced Evidence Analysis

✓ Analysis on URL encoding

- **Most search engines use UTF-8 encoding method.**

  EX) google.com, yahoo.com, bing.com, ask.com, aol.com, excite.com

- **In special case, some site uses multiple encoding method for multiple words in a single URL.**

  ➔ Need for methodology to decide which encoding method is used

http://whdd305.webhard.co.kr/webII/whexp-frame-list.php?loc=%2FGUEST%C6%FA%B4%F5%2FRBR%B0%F8%C5%EB%C0%DA%B7%E1&filename=RBR%EA%B3%B5%ED%86%B5%EC%9E%90%EB%A3%8C

Analysis on URL encoding

Unicode Encoding

- %uHHHH or %26%23<decimal digit>%3B

  EX) %26%2354252%3B%26%2347116%3B%26%2349885%3B ➔ "포렌식" (= forensic)
- Search engine : www.baidu.com

UTF-8 Encoding

- 1~4 byte encoding, 1 byte encoding is same as ASCII

  EX) %ED%8F%AC%EB%A0%8C%EC%8B%9D ➔ "포렌식"(= forensic)
- Most popular encoding method
- Search engine : www.google.co.kr, www.yahoo.co.kr, www.bing.com, www.ask.com

Code Range	UTF-16(UNICODE)	UTF-8	Description
000000-00007F	00000000 0xxxxxxx	**0xxxxxxx**	It is same as ASCII
000080-0007FF	00000xxx xxxxxxxx	**110xxxxx 10xxxxxx**	First byte starts with 110x(= C or D), the rest start with 10
000800-00FFFF	xxxxxxxx xxxxxxxx	**1110xxxx 10xxxxxx 10xxxxxx**	First byte starts with 1110(= E), the rest start with 10
010000-10FFFF	110110yy yyxxxxxx 110111xx xxxxxxxx	**11110zzz 10zzxxxx 10xxxxxx 10xxxxxx**	First byte starts with 11110(yyyy = zzzzz - 1), the rest start wirth 10

Advanced Evidence Analysis

✓ Analysis on URL encoding

– Code Page Encoding

- **In case of korean, 2 byte encoding with EUC-KR code page**

  EX) %C6%F7%B7%BB%BD%C4 ➔ “포렌식”(= forensic)

- **According to code page, same HEX value doesn’t mean same characters**

  ➔ %C6%F7%B7%BB%BD%C4 ➔ EUC-KR ➔ “포렌식”

  ➔ EUC-JP ➔ “匂兄縦”

  ➔ Need for classification of code page according to search engine

- **Search Engine: www.naver.com, www.daum.net, www.nate.com, www.hatena.ne.jp**

Advanced Evidence Analysis

- ✓ **Analysis of user activity**
  - **Difficulty detect the user activity through single piece of URL**

Advanced Evidence Analysis

- ✓ **Analysis of user activity**
  - **The URL include some keyword showing the web site content and use r activity.**

Advanced Evidence Analysis

✓ Analysis of user activity

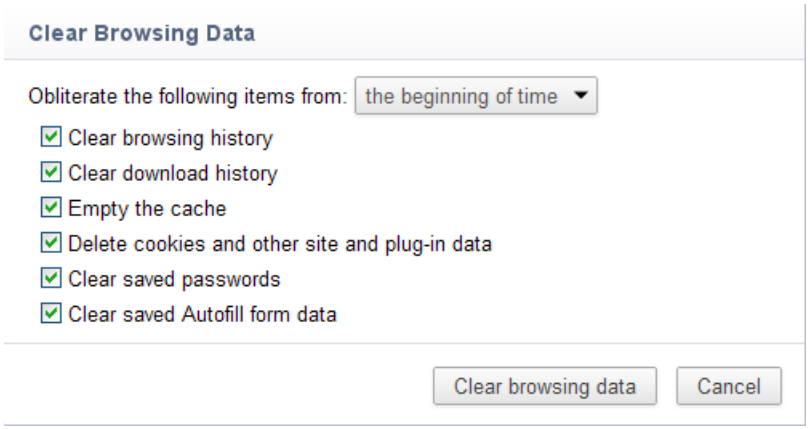
– User Activities in a Web Browser

User Activity	Keyword in URL
Search	Existence of Searched words
Mail	Mail
Blogging	Blog
SNS	Facebook, Twitter …
News	News
Weather	Weather
Shopping	Shopping, Amazon …
Game	Game
Audio-Visual content	Video
Music	Music
Banking	Bank

Advanced Evidence Analysis

✓ Recovery of Deleted Information

- **Most browser provide an erase function for log information.**



Advanced Evidence Analysis

✓ Recovery of deleted information

– Classification of erase function

- **Initialization**

```
00028EF0 | 74 63 6C 69 70 47 19 03  37 01 68 74 74 70 3A 2F | tclipG  7 http:/
00028F00 | 2F 77 77 77 2E 6E 61 76  65 72 2E 63 6F 6D 2F 0E | /www.naver.com/
00028F10 | 81 29 04 82 55 01 68 74  74 70 3A 2F 2F 77 77 77 | ) U http://www
```

```
00028EF0 | 00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00 |
00028F00 | 00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00 |
00028F10 | 00 00 00 00 00 00 00 00  00 00 00 00 00 00 00 00 |
```

- **Deletion log file**

Advanced Evidence Analysis

✓ Recovery of deleted information

– Methods of Erasing Log Information in Five Web Browsers.

Browser	Category	Erasing Method
IE	Cache	Initialization of index.dat file Deletion of Temporary Internet files
	History	Initialization of index.dat file Deleting daily/weekly index.dat files
	Cookie	Initialization of index.dat file Deletion of cookie files
	Download	IE has no download information
Firefox	Cache	Initialization
	History	Initialization
	Cookie	Initialization
	Download	Initialization
Chrome	Cache	Deletion
	History	Initialization
	Cookie	Initialization
	Download	Initialization
Safari	Cache	Initialization
	History	Initialization
	Cookie	Deletion
	Download	Initialization
Opera	Cache	Initialization
	History	Initialization
	Cookie	Initialization
	Download	Initialization

Advanced Evidence Analysis

✓ Recovery of deleted information

– Recovery Method for Deleted Information in Five Web Browsers

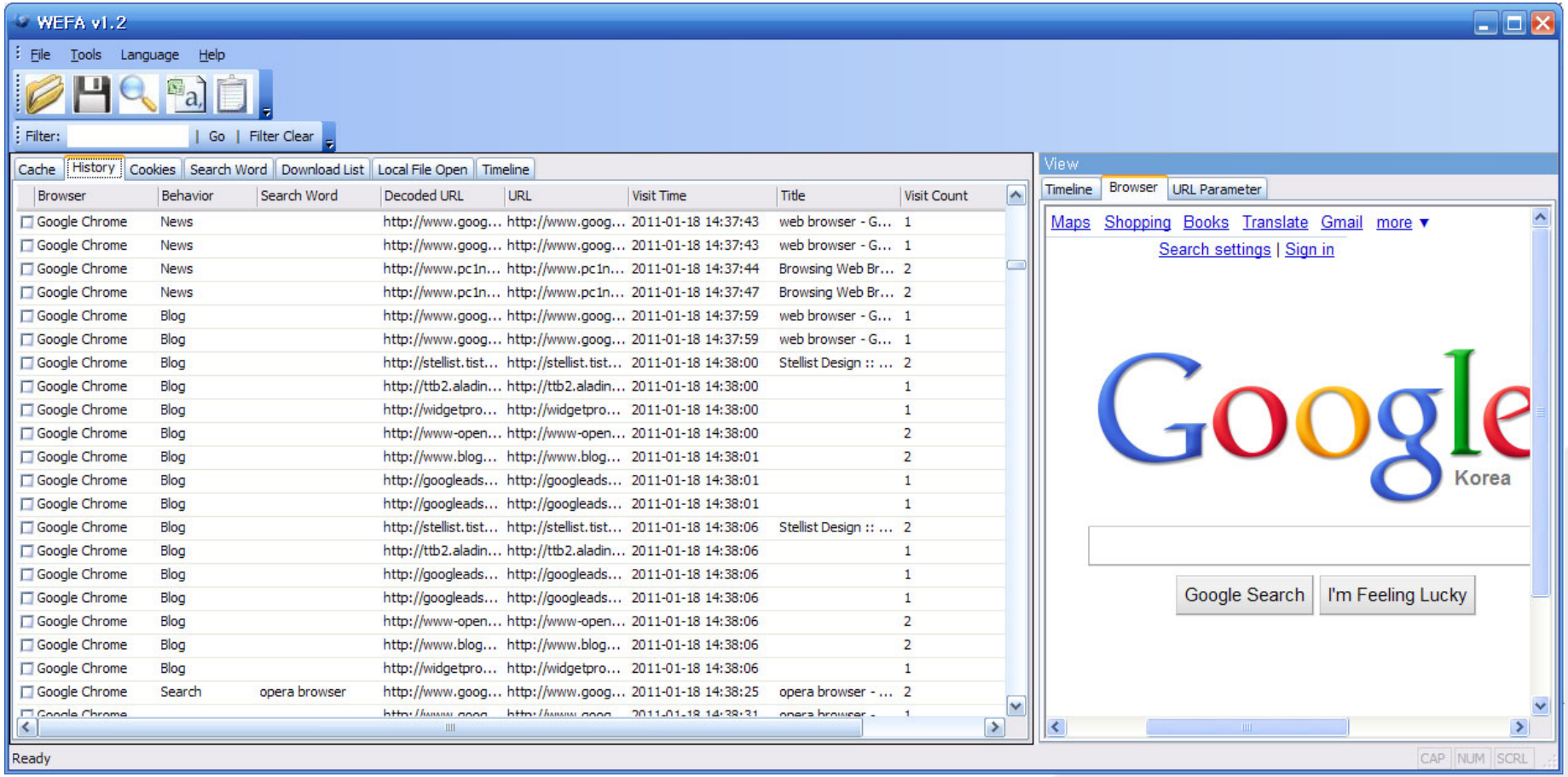
Browser	Category	Recovery Method
IE	Cache	Recovery of temporary Internet files
	History	Recovery of weekly/daily index.dat files Recovery of index.dat file through carving method
	Cookie	Recovery of cookie files
	Download	IE has no download information
Firefox	Cache	N/A
	History	Recovery of session file through carving method
	Cookie	N/A
	Download	N/A
Chrome	Cache	Recovery of cache files
	History	Recovery of monthly history files
	Cookie	N/A
	Download	N/A
Safari	Cache	N/A
	History	Recovery of session files
	Cookie	Recovery of cookie files
	Download	N/A
Opera	Cache	N/A
	History	Recovery of session files
	Cookie	N/A
	Download	N/A

Advanced Evidence Analysis

- ✓ **Recovery of deleted information**
  - – **Consideration**
    - **Time available for recovering deleted information**
    - **Recovered file's creation time and modification time**
      - – **IE : Temporary internet files, Cookie text file**
      - – **Firefox, Safari, Opera : Session file**
    - **Carving Method**
      - – **IE : Daily index.dat**
      - – **Firefox : sessionstore.js**

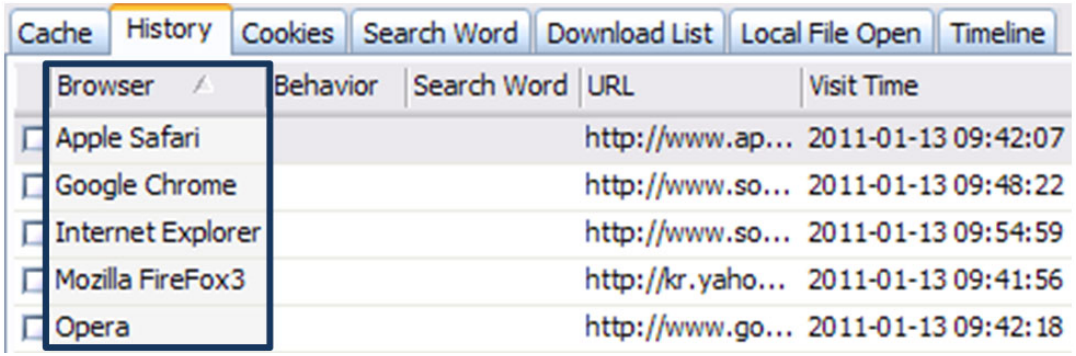
Tool Development

✓ WEFA(Web Browser Forensic Analyzer)



Tool Development

✓ Integrated analysis



✓ Time zone setting

Tool Development

✓ Search word extraction

Cache | History | Cookies | Search Word | Download List | Local File Open | Timeline

Browser	Search Word	URL	Visit Time
Google Chrome	municipal	http://alldic.nate...	2010-10-12 13:11:00
Google Chrome	vulunerability	http://alldic.nate...	2010-10-12 13:11:23
Google Chrome	vulnerability	http://alldic.nate...	2010-10-12 13:11:29
Google Chrome	exploit	http://alldic.nate...	2010-10-12 13:11:41
Google Chrome	malicious	http://alldic.nate...	2010-10-12 13:12:09

✓ URL Decoding

Cache | History | Cookies | Search Word | Download List | Local File Open | Timeline

Browser	Search Word	URL	Visit Time
Mozilla FireFox3	computer	http://www.goog...	2011-01-13 22:12:05
Mozilla FireFox3	ΥπολογιστέςC	http://www.goog...	2011-01-13 22:12:25
Mozilla FireFox3	КомпьютеровD	http://www.goog...	2011-01-13 22:13:15
Mozilla FireFox3	计算机	http://www.goog...	2011-01-13 22:15:45
Mozilla FireFox3	コンピュータ上	http://www.goog...	2011-01-13 22:17:17
Mozilla FireFox3	컴퓨터	http://www.goog...	2011-01-13 22:17:26
Mozilla FireFox3	أجهزة الكمبيوترD	http://www.goog...	2011-01-13 22:18:24

Tool Development

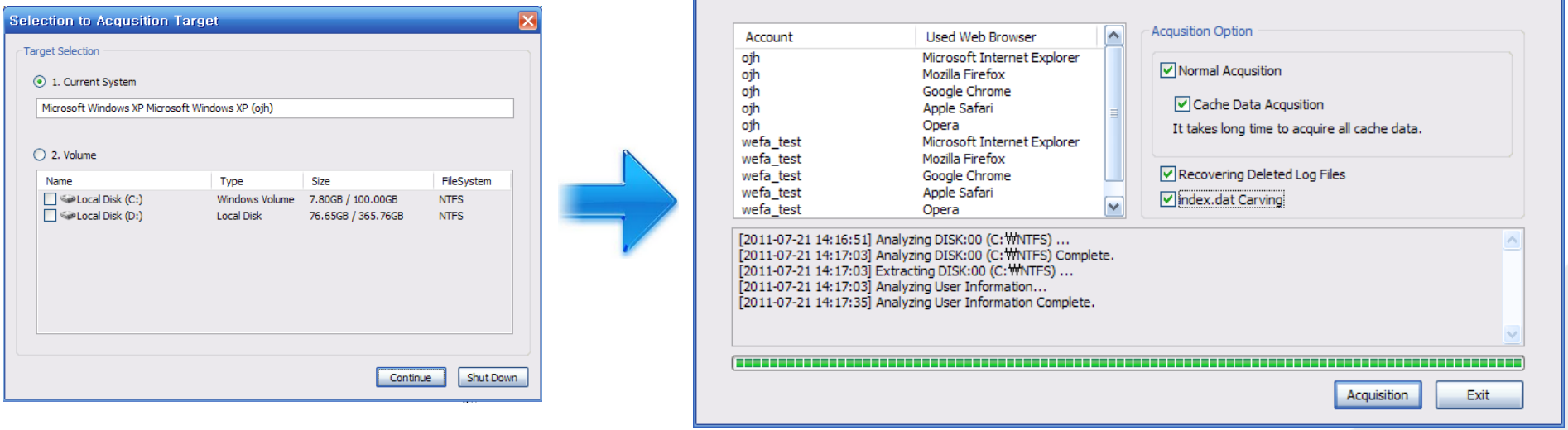
✓ Analysis on user activity

Cache | History | Cookies | Search Word | Download List | Local File Open | Timeline

Browser	Behavior	Search Word	Decoded URL	URL	Visit Time
Internet Explorer	Blog		http://blog.daum...	http://blog.daum...	2011-07-30 23:22:36
Internet Explorer	Blog		http://blog.daum...	http://blog.daum...	2011-07-30 23:22:36
Internet Explorer	Blog		http://widgetpro...	http://widgetpro...	2011-07-30 23:22:36
Internet Explorer	Blog		http://widgetpro...	http://widgetpro...	2011-07-30 23:22:36
Internet Explorer	Blog		http://widgetpro...	http://widgetpro...	2011-07-30 23:22:36
Internet Explorer			http://www.widg...	http://www.widg...	2011-07-30 23:22:40
Internet Explorer			http://kr.koreana...	http://kr.koreana...	2011-07-30 23:23:16
Internet Explorer			http://kr.koreana...	http://kr.koreana...	2011-07-30 23:23:16
Internet Explorer	Blog		http://blog.naver...	http://blog.naver...	2011-07-30 23:23:53
Internet Explorer	Search	esta departure r...	http://blog.naver...	http://blog.naver...	2011-07-30 23:23:53
Internet Explorer	Blog		http://blog.naver...	http://blog.naver...	2011-07-30 23:23:57
Internet Explorer	Blog		http://external.w...	http://external.w...	2011-07-30 23:23:57
Internet Explorer	Search	esta departure r...	http://cafeblog.s...	http://cafeblog.s...	2011-07-30 23:26:38
Internet Explorer	Cafe&Club		http://cafeblog.s...	http://cafeblog.s...	2011-07-30 23:26:44
Internet Explorer	Blog		http://blog.naver...	http://blog.naver...	2011-07-30 23:30:15
Internet Explorer			http://kr.koreana...	http://kr.koreana...	2011-07-30 23:30:17
Internet Explorer	Blog		http://blog.daum...	http://blog.daum...	2011-07-30 23:30:18
Internet Explorer	Blog		http://extrad.egl...	http://extrad.egl...	2011-07-30 23:30:19
Internet Explorer	Blog		http://blog.naver...	http://blog.naver...	2011-07-30 23:30:20
Internet Explorer	Blog		http://blog.naver...	http://blog.naver...	2011-07-30 23:30:24
Internet Explorer	Knowledge		http://kin.naver....	http://kin.naver....	2011-07-30 23:30:25
Internet Explorer	Blog		http://blog.naver...	http://blog.naver...	2011-07-30 23:30:26
Internet Explorer	Blog		http://blog.naver...	http://blog.naver...	2011-07-30 23:30:32
Internet Explorer	Search	ESTA I-94W	http://blog.naver...	http://blog.naver...	2011-07-30 23:30:32

Tool Development

- ✓ Low level acquisition and recovering deleted information



- ✓ Download
  - http://www.4n6tech.com/skin_kr/images/WEFA_v1.2_-_Freeware.zip

Conclusions and Future Research

✓ Advanced Evidence Analysis

- Integrated Analysis
- Time zone Analysis
- Search Word Analysis
- URL Encoding Analysis
- User Activity Analysis
- Recovery of Deleted Information

✓ Future Research

- Linux, Mac Web Browser Forensics

Thank you for attention

blue0226@korea.ac.kr