Web Browser Forensics : Part3

blueangel

blueangel1275@gmail.com

http://blueangel-forensic-note.tistory.com

목차

1. 웹 브라우저 로그 분석 : 심화 과정

- 통합 타임라인 분석
- Time Zone 분석
- 검색어 추출
- URL 인코딩 분석
- 사용자 행위 분류
- 삭제 로그 정보 복구

2. 결론

웹 브라우저 로그 분석 : 심화 과정

- **통합 타임 라인 분석**
- Time Zone 분석
- 검색어 추출
- URL 인코딩 분석
- 사용자 행위 분류
- 삭제 로그 정보 복구

통합 타임라인 분석

- **타임라인 분석?**
  - 웹 브라우저의 사용 내역을 일련의 시간 흐름으로 재구성
  - 사용자의 사이트간 이동경로를 파악, 전체적인 사용자 행위를 유추

- **통합 타임라인 분석의 필요성**
  - 한 명의 사용자가 여러 브라우저를 동시에 사용하는 경우가 많아짐
    ➔ 여러 브라우저의 로그를 하나의 타임라인 상에서 분석해 주는 도구 필요~!!!
  - **각 로그의 시간 정보를 기준으로 통합**

통합 타임라인 분석

- **5대 웹 브라우저 모두 서로 다른 시간 포멧 사용**
  - **하나의 공통된 포멧으로 통일할 필요성이 있음**

웹 브라우저	시간 정보 포멧
Internet Explorer	FILETIME: 100-nanosecond (10^{-9}) Since January 1, 1601 00:00:00 (UTC+0)
Firefox	PRTime: microsecond(10^{-6}) Since January 1, 1970 00:00:00 (UTC+0)
Chrome	WEBKIT Time: microsecond(10^{-6}) Since January 1, 1601 00:00:00 (UTC+0)
Safari	CFAbsoluteTime: second Since January 1, 2001 00:00:00 (UTC+0)
Opera	UNIX Time: second Since January 1, 1970 00:00:00 (UTC+0)

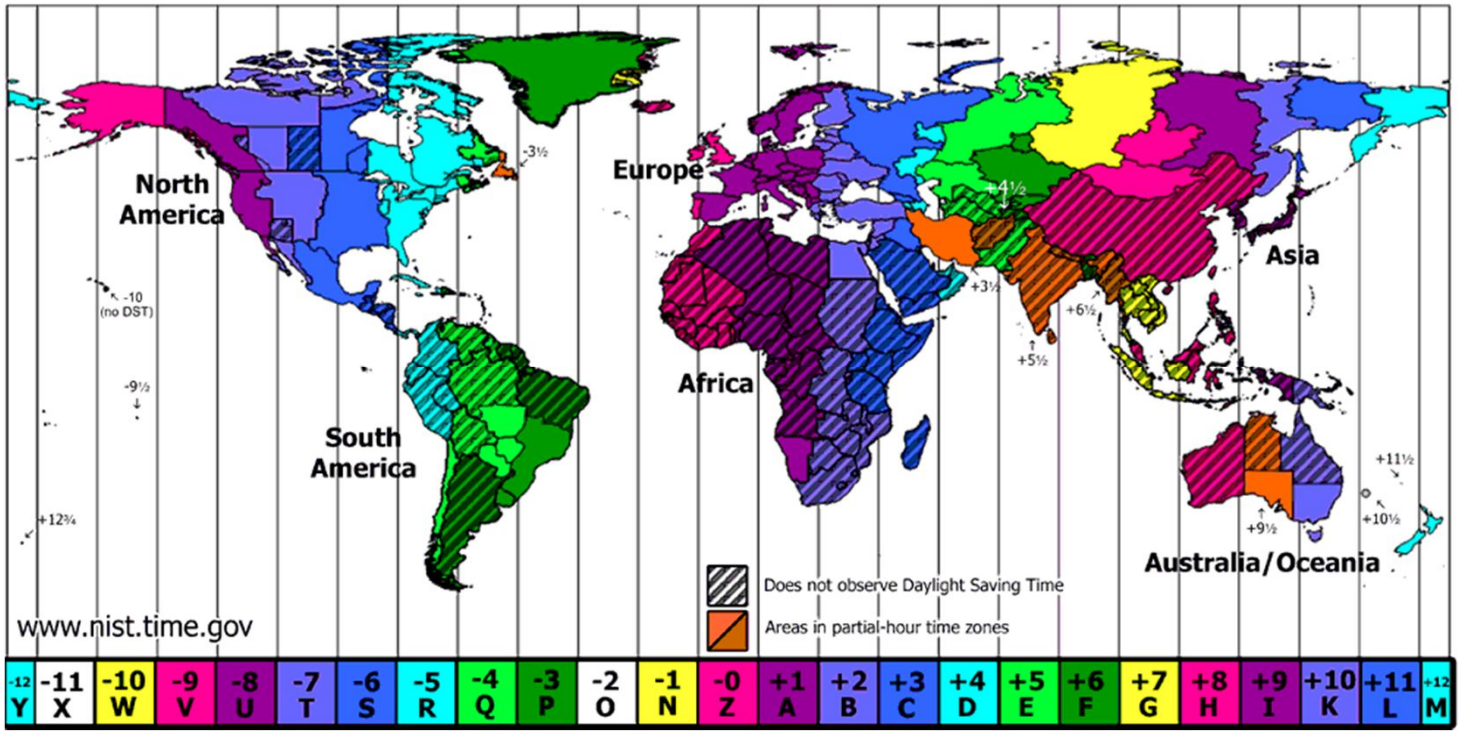
웹 브라우저 로그 분석 : 심화 과정

- 통합 타임 라인 분석
- **Time Zone 분석**
- 검색어 추출
- URL 인코딩 분석
- 사용자 행위 분류
- 삭제 로그 정보 복구

타임존 분석

- **5대 웹 브라우저 모두 UTC+0을 기준으로 시간 정보 저장**
  - 로그 수집 장소의 지역 시간으로 변경할 필요성이 있음
    ➔ 분석 도구의 **타임존 변환 기능** 필요~!!!

  - 예외적으로 지역 시간으로 저장되는 경우가 있음
    ➔ IE 주간 히스토리 정보의 접근 시간



웹 브라우저 로그 분석 : 심화 과정

- 통합 타임 라인 분석
- Time Zone 분석
- **검색어 추출**
- URL 인코딩 분석
- 사용자 행위 분류
- 삭제 로그 정보 복구

검색어 추출

- **검색어?**
  - 사용자의 의도와 관심사를 가장 명확하게 알 수 있는 키워드
  - 웹 브라우저 포렌식 조사 과정에서 가장 중요한 정보

- **일반적으로 검색 엔진에 입력한 검색어는 Cache, History 로그의 URL에 기록됨**

- **기본적인 URL 구조**

프로토콜://	도메인	/	경로(/../../)	페이지 파일	?	변수=값	...

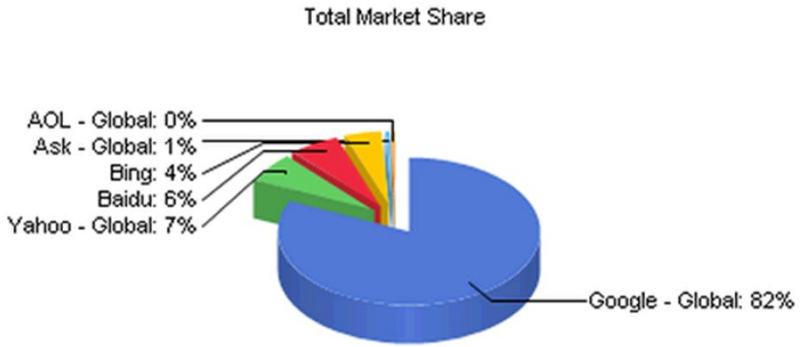
- **구글 검색 엔진에서 “forensic” 키워드 검색시, 로그에 저장되는 URL**

http://www.google.co.kr/#sclient=psy&hl=ko&newwindow=1&source=hp&**q=forensic**&aq=f&aqi=g5&aql=&oq=&pbx=1&fp=a03afb05b2691392&biw=1280&bih=939

도메인	변수	값	변수	값	변수	값	변수	값	변수	값
Google.co.kr	#sclient	psy	hl	ko	newwindow	1	source	hp	q	forensic

검색어 추출

- 전세계 검색 엔진 점유율 (출처 : NetMarketShare, 2012년 6월)



Search Engine	Total Market Share
Google - Global	80.77%
Yahoo - Global	6.57%
Baidu	6.06%
Bing	4.43%
Ask - Global	0.53%
AOL - Global	0.35%
Excite - Global	0.03%
AltaVista - Global	0.01%
Lycos - Global	0.01%

검색어 추출

- 검색어 추출을 위한 각 검색 엔질 별 시그니처와 검색어 위치

검색 엔진	도메인	시그니처	검색어 위치
Google	google.com	sclient	변수 “q” 의 값
Yahoo	search.yahoo.com	/search	변수 “p” 의 값
Baidu	baidu.com	/s	변수 “wd” 의 값
Bing	bing.com	/search	변수 “q” 의 값
Ask	ask.com	/web	변수 “q” 의 값
AOL	search.aol.com	/search/	변수 “q” 의 값
Excite	msxml.excite.com	/results/	경로 “/Web/” 뒤에 위치
Lycos	Search.lycos.com		변수 “query” 의 값
Alta vista	altavista.com	/search	변수 “p” 의 값
MSN	bing.com	/search	변수 “q” 의 값
Naver	naver.com	/search	변수 “query” 의 값
Daum	daum.net	/search	변수 “q” 의 값
Nate	nate.com	/search	변수 “q” 의 값

웹 브라우저 로그 분석 : 심화 과정

- 통합 타임 라인 분석
- Time Zone 분석
- 검색어 추출
- **URL 인코딩 분석**
- 사용자 행위 분류
- 삭제 로그 정보 복구

URL 인코딩 분석

- URL 에서 영어가 아닌 문자는 모두 인코딩됨 ➔ URL 디코딩 필요~!!!

URL 인코딩 분석

- **기본적인 URL 인코딩 방식**
  - 헥사코드로 표현 한 후, 1바이트단위로 '%' 문자를 앞부분에 추가
    ➔ %ED%8F%AC%EB%A0%8C%EC%8B%9D

- **URL 인코딩 방식 분류**
  - UTF-8 인코딩
    - ✓ 1~4바이트 까지 사용, 1바이트 까지는 ASCII와 동일
      EX) %ED%8F%AC%EB%A0%8C%EC%8B%9D ➔ "포렌식"
    - ✓ 전세계에서 가장 많이 사용하는 인코딩 형식
    - ✓ 검색 엔진 : www.google.co.kr, www.yahoo.co.kr, www.bing.com, www.ask.com …

코드 범위	UTF-16(UNICODE)표현	UTF-8 표현	설명
000000-00007F	00000000 0xxxxxxx	**0xxxxxxx**	ASCII와 동일한 범위
000080-0007FF	00000xxx xxxxxxxx	**110xxxxx 10xxxxxx**	첫 바이트는 110으로 시작하고, 나머지 바이트들은 10으로 시작함
000800-00FFFF	xxxxxxxx xxxxxxxx	**1110xxxx 10xxxxxx 10xxxxxx**	첫 바이트는 1110으로 시작하고, 나머지 바이트들은 10으로 시작함
010000-10FFFF	110110yy yyxxxxxx 110111xx xxxxxxxx	**11110zzz 10zzxxxx 10xxxxxx 10xxxxxx**	UTF-16 서러게이트 쌍 영역 (yyyy = zzzzz - 1). UTF-8로 표시된 비트 패턴은 실제 코드 포인트와 동일하다.

URL 인코딩 분석

▪ URL 인코딩 방식 분류(계속)

- Unicode 인코딩
  - ✓ Unicode의 2바이트를 16진수 혹은 10진수 형식으로 인코딩
    - ➔ **%uHHHH** or **%26%23<십진수>%3B** 형식으로 인코딩 됨
    - EX) %26%2354252%3B%26%2347116%3B%26%2349885%3B ➔ "포렌식"
  - ✓ 검색 엔진 : www.baidu.com

- 코드 페이지 인코딩
  - ✓ 각 나라의 코드 페이지에 따라 인코딩하는 방식
  - ✓ 한국어의 경우, EUC-KR(한글 완성형) 코드표에 따라 2바이트 단위로 인코딩 됨
    - EX) %C6%F7%B7%BB%BD%C4 ➔ "포렌식"
  - ✓ 같은 HEX 값이라도 코드 페이지에 따라 의미하는 문자가 달라짐
    - ➔ %C6%F7%B7%BB%BD%C4 ➔ EUC-KR ➔ "포렌식"
    - ➔ EUC-JP ➔ "匂兄縦"
    - ➔ 사이트 혹은 국가별 구분 필요
  - ✓ 검색 엔진 : www.nate.com, www.paran.com

웹 브라우저 로그 분석 : 심화 과정

- 통합 타임 라인 분석
- Time Zone 분석
- 검색어 추출
- URL 인코딩 분석
- **사용자 행위 분류**
- 삭제 로그 정보 복구

사용자 행위 분류

- **필요성**
  - URL 정보만으로는 사용자 행위를 바로 파악이 어려움

    ➔ 웹 브라우저를 통해 직접 웹 사이트를 방문 필요 ➔ 시간 소모~

  - 분석 효율성 증가 ➔ 분석 시간 단축
    - ✓ 수 많은 로그 기록에서 사용자 행위 흐름을 한 눈에 파악 가능
    - ✓ 조사자가 찾고자 하는 사용자 행위를 빠르게 검색

사용자 행위 분류

- **분류 방법**
  - History URL은 웹 사이트 제작자에 의해 결정됨
    ➔ URL의 도메인과 Path에는 해당 웹 페이지의 특징과 역할을 나타내는 키워드 존재
  - 키워드를 기반으로 사용자 행위 분류
    - ✓ 일반적인 키워드 : mail, lecture, map, news
    - ✓ 특수 키워드 : facebook, twitter, ➔ DB에 저장된 시그니처 사용

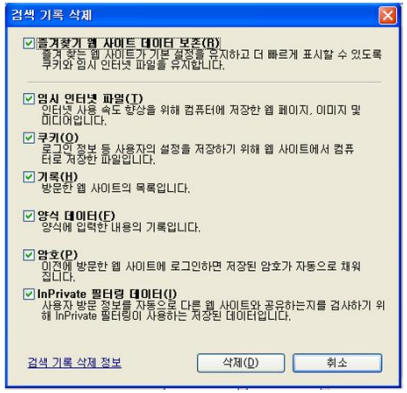
사용자 행위		일반 키워드 및 분류 방법
정보 획득	검색	검색어 추출이 가능하면 검색 행위로 분류
	위키피디아	wikipedia
	문서 보기	.doc(x), .ppt(x), .xls(x), .pdf, .txt …
	블로그	blog
	교육	lecture, study, learn
	지도	map
	뉴스	news
	날씨	weather
	사전	dic, dictionary
다운로드		ftp://, download
구입	쇼핑	shop, store
	예약	reservation, ticket
인간관계	SNS	DB 에 저장된 시그니처 사용(facebook.com, twitter.com …)
	커뮤니티	forum, cafe, club, group, community
	채팅	chat
업무	메일	mail, email, e-mail
	일정	calendar
	클라우드	DB 에 저장된 시그니처 사용(docs.google.com, zoho, aws.amazon, racksace …)
	은행업무	bank
	신청	apply, application
	등록	register
엔터테이먼트		game, movie, music, cartoon, sport, radio, adult, porn, sex

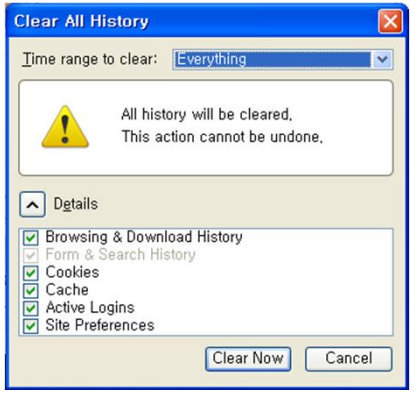
웹 브라우저 로그 분석 : 심화 과정

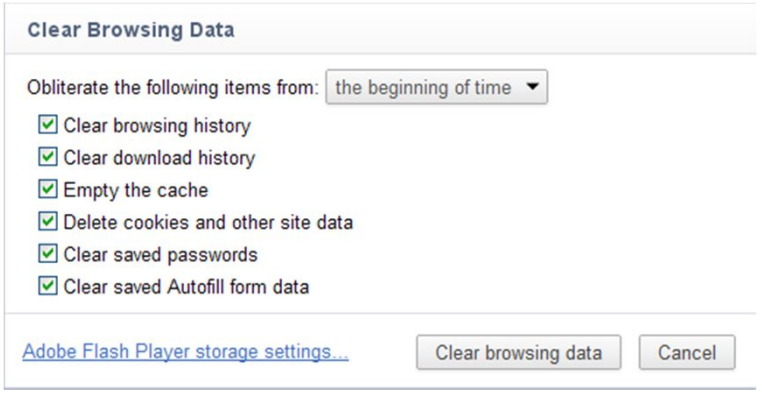
- 통합 타임 라인 분석
- Time Zone 분석
- 검색어 추출
- URL 인코딩 분석
- 사용자 행위 분류
- **삭제 로그 정보 복구**

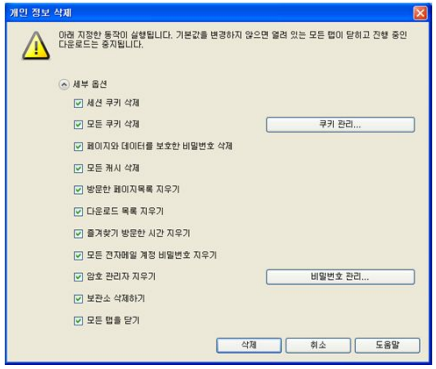
삭제된 로그 정보 복구

- **대부분의 웹 브라우저들은 로그 삭제 기능을 가지고 있음**
  ➔ 웹 브라우저 포렌식 조사 과정을 어렵게 만듬



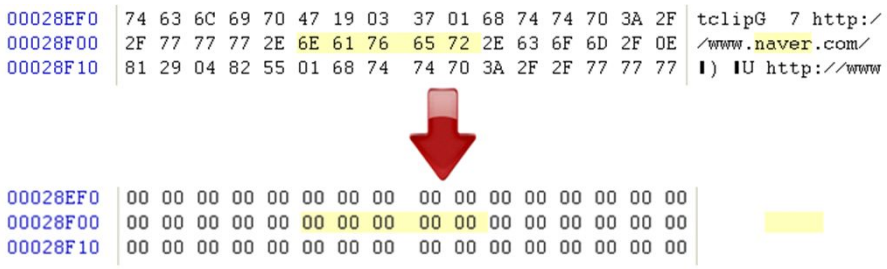






삭제된 로그 정보 복구

- 사용자 정보 삭제 기능 분류
  - 초기화 : 해당 파일의 내용을 초기화 (0으로 혹은 파일 초기 상태로)



  - 파일 삭제 : 해당 파일을 파일 시스템에서 삭제

- 삭제 기능 분류에 따른 복구 가능 여부
  - 초기화 : 복구 불가능
    - ✓ 유사 정보를 통한 복구
      EX) History 정보가 초기화 되도 Session 정보가 삭제되는 경우가 있음
      ➔ 복구한 세션 파일의 정보를 통해 History 정보의 일부분을 복구 가능

  - 파일 삭제
    - ✓ 파일 시스템 관점에서 삭제된 파일을 복구한 후, 삭제된 정보 추출
    - ✓ 카빙 기법을 통해 삭제된 파일 복구 가능

삭제된 로그 정보 복구

- 각 브라우저 별 로그 정보 삭제 방식

웹 브라우저	로그 정보	삭제 방식
IE	Cache	index.dat 파일의 데이터 초기화 임시 인터넷 파일 삭제
	History	index.dat 파일의 데이터 초기화 일간/주간 index.dat 파일 삭제
	Cookie	index.dat 파일의 데이터 초기화 Cookie 텍스트 파일 삭제
	Download	IE 8 이하는 다운로드 정보 없음 IE 9 다운로드 정보의 레코드는 삭제되지 않음(비활성화)
Firefox	Cache	초기화
	History	초기화
	Cookie	초기화
	Download	초기화
Chrome	Cache	파일 삭제
	History	초기화
	Cookie	초기화
	Download	초기화
Safari	Cache	초기화
	History	초기화
	Cookie	파일 삭제
	Download	초기화
Opera	Cache	초기화
	History	초기화
	Cookie	초기화
	Download	초기화

삭제된 로그 정보 복구

- 각 브라우저 별 삭제된 로그 정보에 대한 복구 방식

웹 브라우저	로그 정보	Recovery Method
IE	Cache	삭제된 임시 인터넷 파일 복구
	History	삭제된 일간/주간 index.dat 파일 복구 비할당 영역에서 일간/주간 index.dat 파일 카빙
	Cookie	삭제된 Cookie 텍스트 파일 복구
	Download	IE 9 다운로드 정보의 레코드는 삭제되지 않음(비활성화)
Firefox	Cache	N/A
	History	삭제된 Session 파일 복구 비할당 영역에서 session 파일 카빙
	Cookie	N/A
	Download	N/A
Chrome	Cache	삭제된 Cache 파일 복구
	History	삭제된 월간 History 파일 복구
	Cookie	N/A
	Download	N/A
Safari	Cache	N/A
	History	삭제된 Session 파일 복구
	Cookie	삭제된 Cookie 파일 복구
	Download	N/A
Opera	Cache	N/A
	History	삭제된 Session 파일 복구
	Cookie	N/A
	Download	N/A

결 론

결론

- **단순히 로그 파일 파싱이 전부가 아니다~!!**
- **웹 브라우저 로그 분석에서 고려해야 할 사항**
  - 여러 웹 브라우저에 대한 통합 타임라인 분석
  - Time Zone 적용
  - 검색어 추출
  - URL 인코딩
  - 사용자 행위 분류
  - 삭제된 로그 정보에 대한 복구
- **분석 시, 위 사항을 고려 및 지원해주는 도구 사용**

질문 및 답변