『国内のフォレンジック』

2005年10月18日15:00 - 16:20

NetAgent

Hideaki Ihara

講師紹介

- 伊原秀明（ihara@netagent.co.jp）
  ネットエージェント株式会社 取締役
- Windowsに対する不正アクセスとその対策方法、“不正アクセス調査”や“コンピュータ・フォレンジック”などを専門に扱う。
- Microsoft MVP (Windows Security)
- 日記（http://d.hatena.ne.jp/hideakii/）

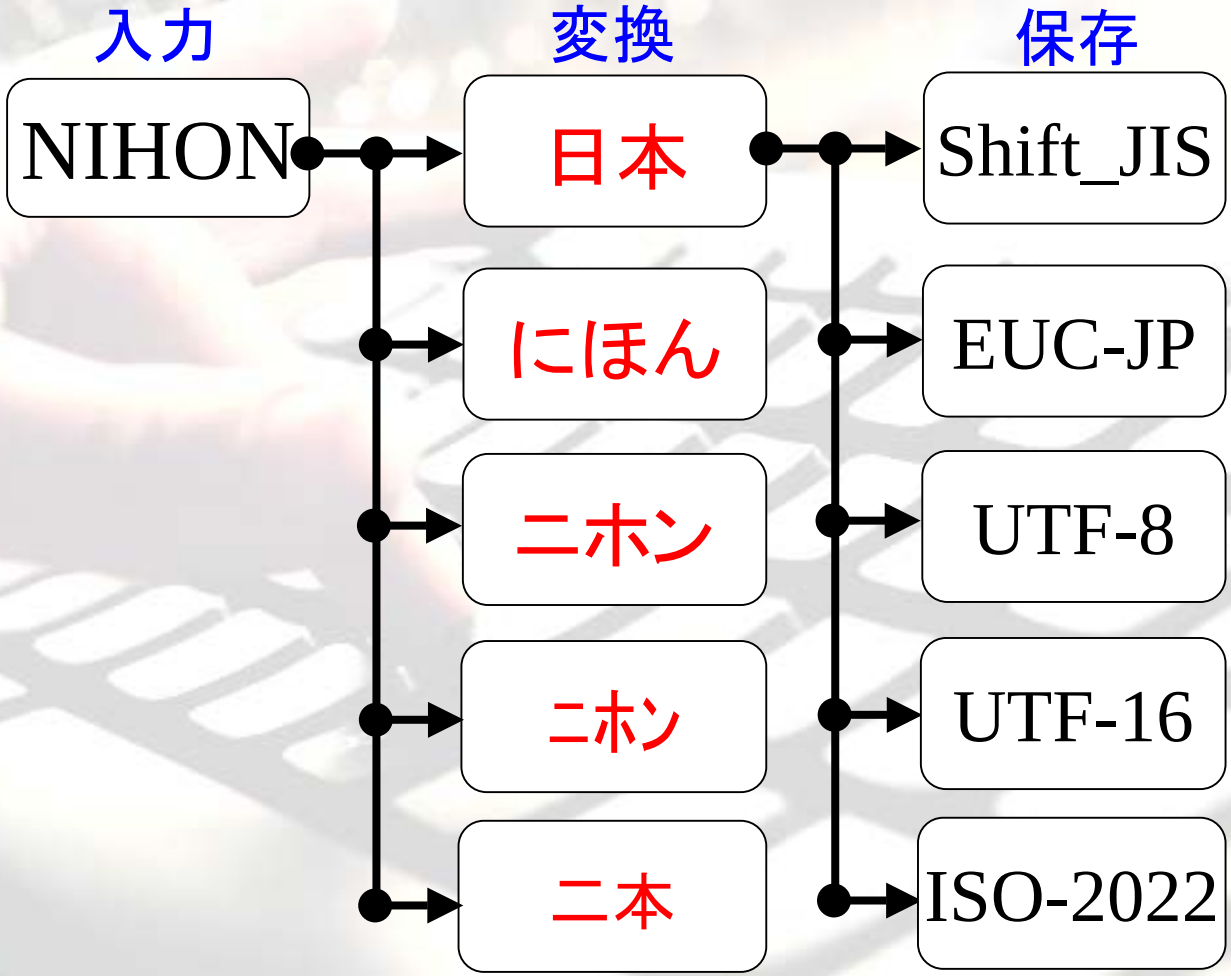
セッションの内容

- フォレンジック調査において、「日本語文字列」を検索することは必須の作業となります。しかし、フォレンジック調査に使用されるツールの多くは海外の製品であり、必ずしも十分に日本語を扱えるわけではありません。
- また、日本では文字コードを利用したアンチ・フォレンジック手法についての研究も行われており、調査員にとって日本語と文字コードは避けて通ることができない大きな壁となっています。本セッションでは、現在一般的に使われている調査ツールで、調査対象として「日本語」を扱う際に注意しなければならない点や、今後の技術的課題について主なテーマとして扱います。

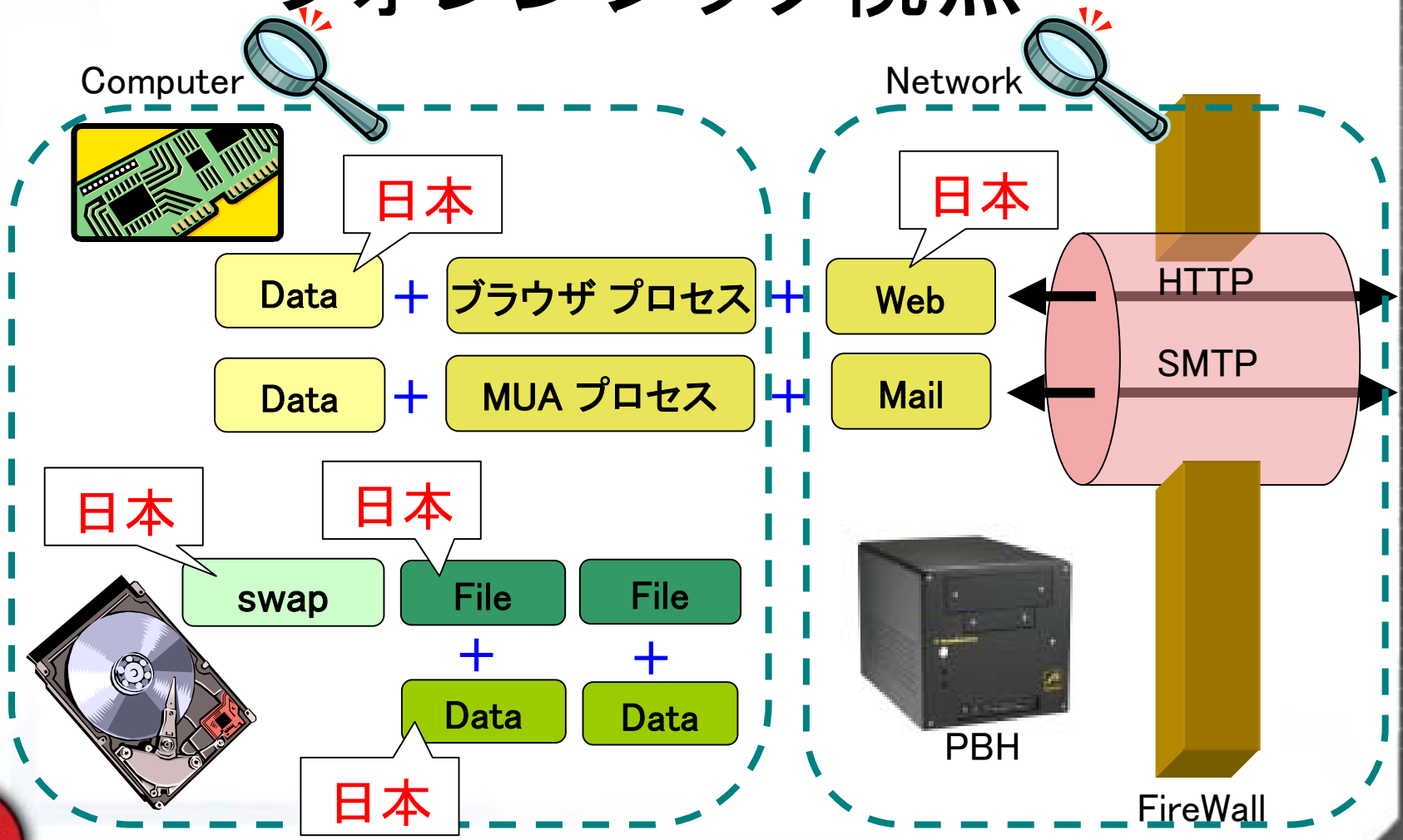
日本語の調査方法

- 目視
- 検索
- 文字列抽出
- インデックス化

『日本』

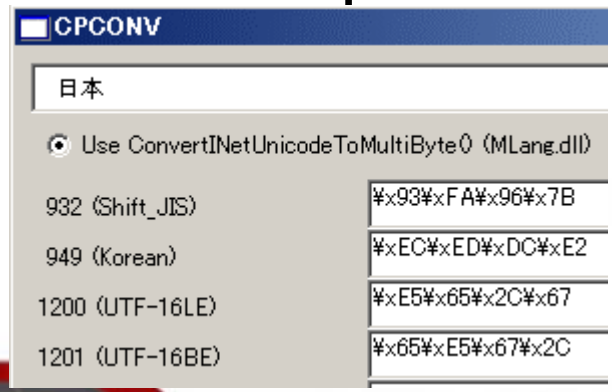

入力
変換
保存
NIHON
日本
にほん
ニホン
ニホン
二本
Shift_JIS
EUC-JP
UTF-8
UTF-16
ISO-2022

フォレンジック視点


Computer
Network
日本
Data
＋
ブラウザ プロセス
Data
＋
MUA プロセス
日本
swap
日本
File
File
＋
＋
Data
Data
日本
日本
Web
Mail
HTTP
SMTP
PBH
FireWall

CPCONV

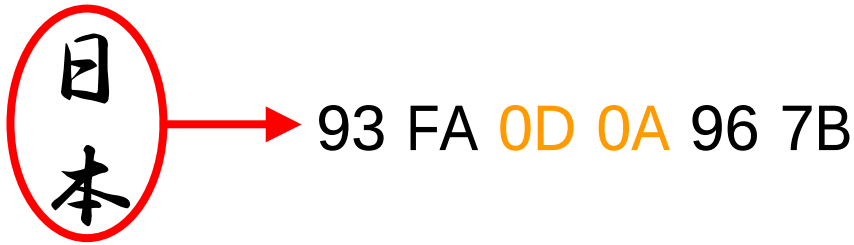
- 入力された文字列を、Windowsのコードページ毎に16進形式で表示
- 16進数で文字列検索を行う際に活躍！
- CPCONV 0.8.1
  https://www.port139.co.jp/forensics/cpconv/



Special thanks to umq.

文字列検索の障害/妨害

- 16進形式のパターン検索は、文字列の途中にゴミが混ざると検索に失敗する
- 意図的に検索を妨害することも可能
- 例）改行コードが存在



- 回避策（EnCaseのgrepオプションを使用）
  日[¥x0d¥x0a]*本

文字列検索の障害/妨害

- Unicode 制御文字の影響
  （1）見えない文字
  （2）見た目が同じ
  （3）方向

U+FEFF; ZERO WIDTH NO-BREAK SPACE
U+200B; ZERO WIDTH SPACE
U+200C; ZERO WIDTH NON-JOINER
U+200D; ZERO WIDTH JOINER
U+202E; RIGHT-TO-LEFT OVERRIDE
U+202C; POP DIRECTIONAL FORMATTING

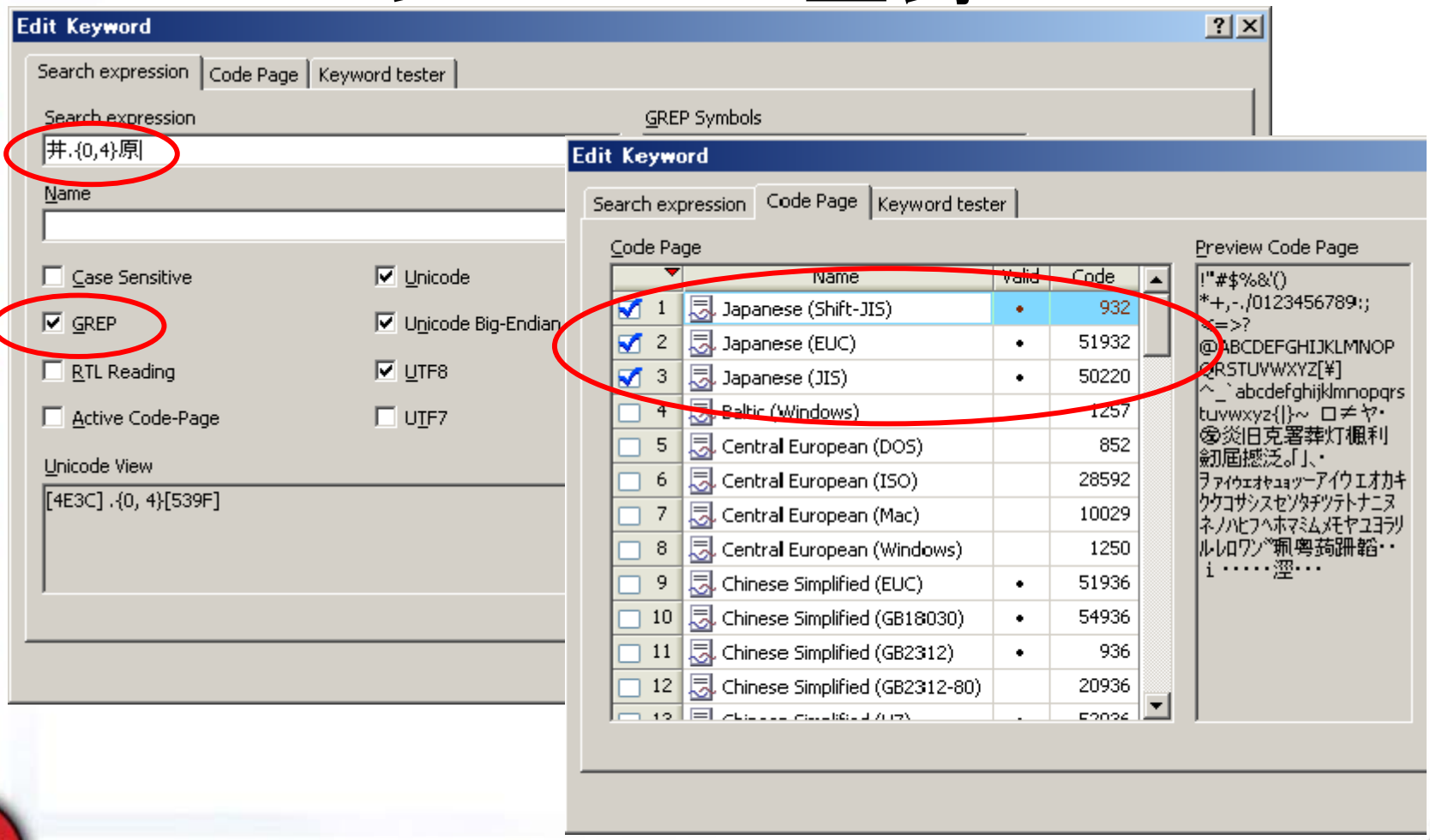
参照URL:
http://www.fileformat.info/info/unicode/
http://www.microsoft.com/windows2000/ja/professional/help/lang_unicode_control_characters.htm

文字列検索の障害/妨害

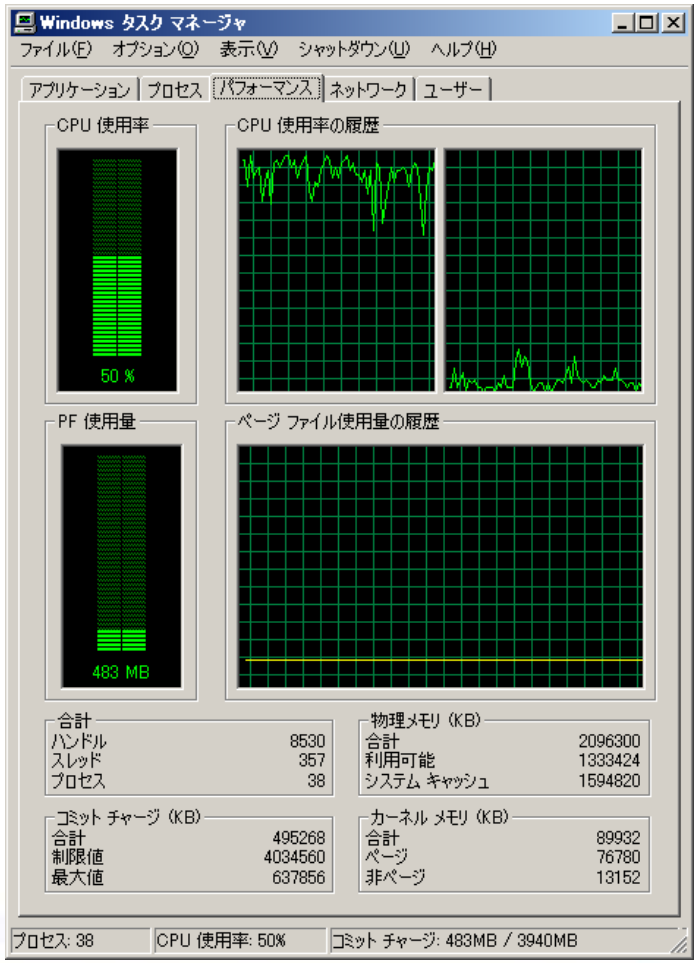
- 回避策（EnCaseのgrepオプションを使用）
  例）日本の間に0~4文字、何かが挟まる
  日.{0,4}本
- 問題点（誤検知の増加・予想が困難）
  u+65E5 u+FFFF u+FFFFu+FFFFu+FFFF u+672C

参照URL：
http://d.hatena.ne.jp/hasegawayosuke/20041121#p3
http://d.hatena.ne.jp/hasegawayosuke/20050106#p1

キーワード登録


Edit Keyword
Search expression
Code Page
Keyword tester
Search expression
井.{0,4}原
GREP Symbols
Name
Case Sensitive
Unicode
GREP
Unicode Big-Endian
RTL Reading
UTF8
Active Code-Page
UTF7
Unicode View
[4E3C] .{0, 4}[539F]
Edit Keyword
Search expression
Code Page
Keyword tester
Code Page
Name
Valid
Code
1 Japanese (Shift-JIS) • 932
2 Japanese (EUC) • 51932
3 Japanese (JIS) • 50220
4 Baltic (Windows) 1257
5 Central European (DOS) 852
6 Central European (ISO) 28592
7 Central European (Mac) 10029
8 Central European (Windows) 1250
9 Chinese Simplified (EUC) • 51936
10 Chinese Simplified (GB18030) • 54936
11 Chinese Simplified (GB2312) • 936
12 Chinese Simplified (GB2312-80) 20936
Preview Code Page

検索実行中



CPU:PentiumD 820
Memory:2GB
HDD:SATA, 7200rpm

検索結果

Searching
Status: Completed
Start: 09/27/05 08:40:31午後
Stop: 09/28/05 06:30:57午前
Time: 9:50:00
Files: 267,808
Search Hits: 843
Added Search Hits: 843

Shift_JIS : 50件
UTF-16LE : 66件
UTF-16BE : 59件
ISO-2022-JP:469件
EUC-JP:199件

単純な検索より
時間がかかる！

文字列抽出

- Stringsコマンドを利用して文字列を抽出
  例）The Sleuth Kit のキーワード検索
- 未使用領域（Unallocated Clusters）、Slack スペース、メモリダンプの調査

参照URL：
sstrings and Unicode Searching
http://www.sleuthkit.org/informer/sleuthkit-informer-16.html
File Name Searching In Autopsy
http://www.sleuthkit.org/informer/sleuthkit-informer-15.html

文字列抽出の利点

- 検索では検出が困難なケースへも対応が可能となる
- 偽装への対応
  例）HKDFの ini ファイル

```
[H<<<idden T>>>a/"ble]
>h"xdef"*
r|c<md¥.ex<e::
```

「Hidden」を検索で発見するのは困難

istrings

- 日本語および Unicode 対応の strings
- istringsで抽出した日本語文字列に対して、別のツールであいまい検索が可能に

例）複数文字コードの抽出結果をUTF−16LEへ変換し結合

```
@echo off
istrings -i EUC-JP -f -p -c $* | wiconv -f 51932 -t 1200 >> log.txt
istrings -i CP932  -f -p -c $* | wiconv -f   932 -t 1200 >> log.txt
istrings -i UTF-16LE -f -p -c $* >> log.txt
```

Special thanks to HASEGAWA Yosuke .

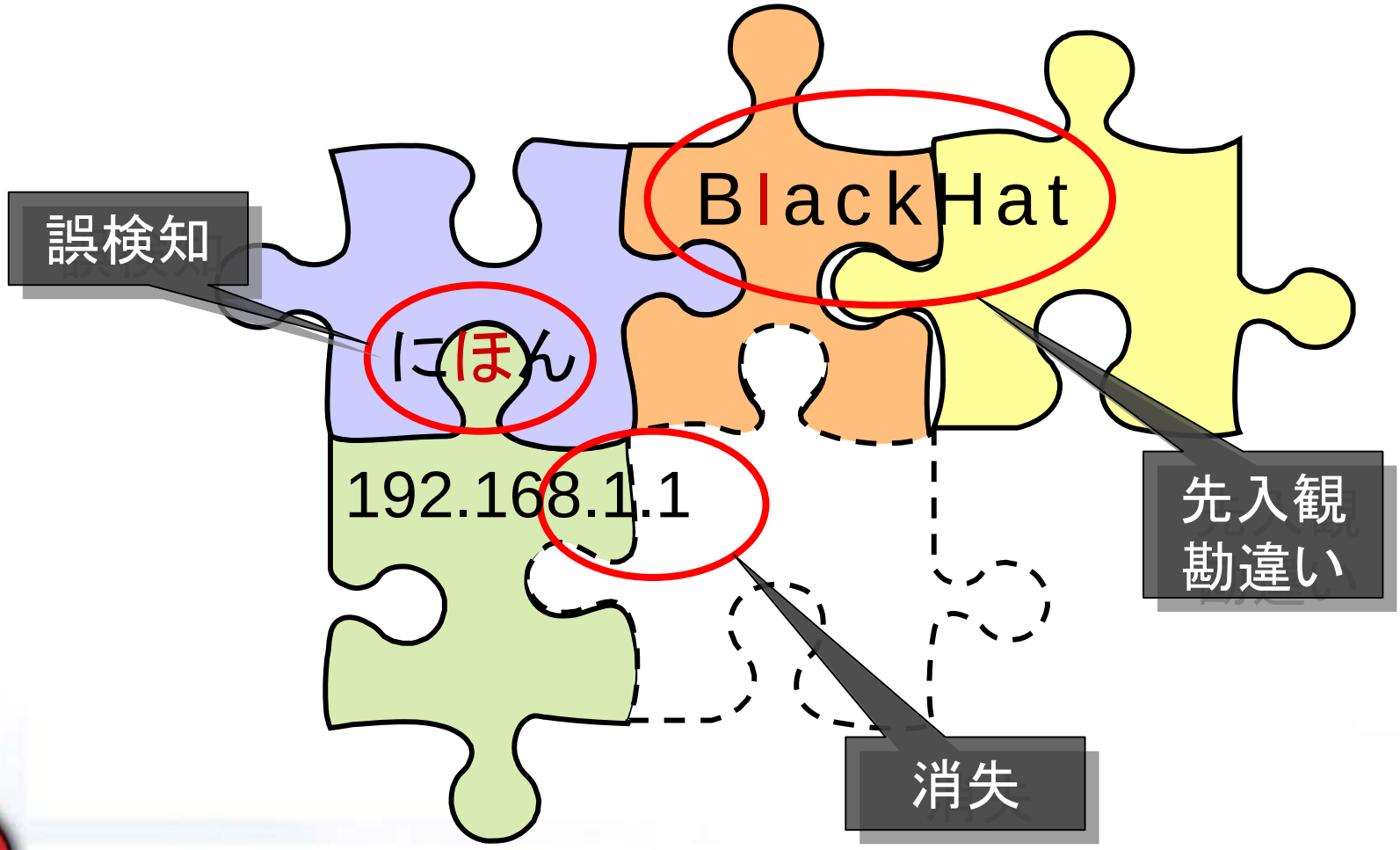
インデックス化

- 文字列をインデックス化することで高速な検索が可能
- 先頭文字の入力によるジャンプ機能
- 対応製品
  FTK, dtSearch, Paraben Text Searcher
- 米国では電子メールの調査によく利用？

独自ファイル形式への対応

- アプリケーション独自の保存形式で記録されており、そのまま検索・文字列抽出しても一致しないファイルフォーマットへの対応
- 形式変換してから検索・文字列抽出する
  例）画像データに書かれた文字列

断片(パズル)の調査



今後の課題

- 調査用ツールの日本語対応
- 検索結果の絞込み（あいまい検索）
- Unicode 正規化へ対応した検索
- 検索・インデックス作成時間の短縮
- 対ローテク・アンチ・フォレンジック手法

参考資料

- 文字コード超研究
  深沢 千尋 (著), ラトルズ ; ISBN: 4899770510
- Unicode標準入門
  トニー グラハム (著), 翔泳社 ; ISBN: 4798100307

日本語対応可能な調査ツール

- EnCase
  http://www.encase.jp/
- Forensic Toolkit（FTK)
  http://www.ubic.co.jp/FTK.htm
- The Sleuth Kit（＋UTF-8 Patch）
  http://www.sleuthkit.org/
  http://www.t-dori.net/

istrings関連URL

- istrings, jstrings, wiconv
  http://openmya.hacker.jp/hasegawa/
- istringsにみんなで色々くっつけよう
  http://openmya.hacker.jp/hiki/
- セキュリティアカデミー勉強会
  http://d.hatena.ne.jp/hasegawayosuke/20050710#1121007299
- umqさんによるバグ報告
  http://d.hatena.ne.jp/hasegawayosuke/20050714#1121330442
- istrings 0.2 を近いうちに出します
  http://d.hatena.ne.jp/hasegawayosuke/20050715#1121408744
- istrings 0.2 リリース
  http://d.hatena.ne.jp/hasegawayosuke/20050717#1121533101
- フォレンジックにおける文字列抽出と検索
  http://forensics.sakura.ne.jp/PPT/20050709-cakeoff-ihara.ppt
- りーふきっと発表会
  http://d.hatena.ne.jp/hideakii/20050709

参考資料：Windowsにおけるフォントセットとサポートされるコードページ
http://www.microsoft.com/japan/office/ork/three/inte03.asp の表を引用

Office XP に付属のフォント

フォント (ファイル)	コード ページ	サポートされる言語
Arial Unicode MS (Arialuni.ttf)	すべて	すべて
Batang (Batang.ttf)	250、1251、1252、1253、1254、1257、949	ほとんどのヨーロッパ言語、韓国語
PmingLiu (PMingliu.ttf)	932, 936, 950	英語、簡体字中国語、繁体字中国語、日本語
MS 明朝 (Msmincho.ttf)	1250、1251、1252、1253、1254、1257、932	ほとんどのヨーロッパ言語、日本語
SimSun (Simsun.ttf)	936	英語、簡体字中国語、繁体字中国語
Georgian and Armenian Font (Sylfaen.ttf)	1250、1251、1252、1253、1254、1257、Unicode	ほとんどのヨーロッパ言語、グルジア語、アルメニア語
Hindi Font (Mangal.ttf)	(Unicode)	ヒンディー語
Tamil Font (Latha.ttf)	(Unicode)	タミール語

参考資料：Windows環境におけるコードページ

http://msdn.microsoft.com/library/en-us/intl/unicode_81rn.asp より抜粋

Code-Page Identifiers

Identifier	Name
932	ANSI/OEM - Japanese, Shift-JIS
1200	Unicode UCS-2 Little-Endian (BMP of ISO 10646)
1201	Unicode UCS-2 Big-Endian
20932	JIS X 0208-1990 & 0121-1990
50220	ISO 2022 Japanese with no halfwidth Katakana
50221	ISO 2022 Japanese with halfwidth Katakana
50222	ISO 2022 Japanese JIS X 0201-1989
51932	EUC - Japanese
65000	Unicode UTF-7
65001	Unicode UTF-8