Finding and Querying on Document Metadata

[redacted]
Booz|Allen|Hamilton
Sigint Development Support / SIGINT Technical Analysis (SDS/STA)
April 2009

Document Metadata Agenda

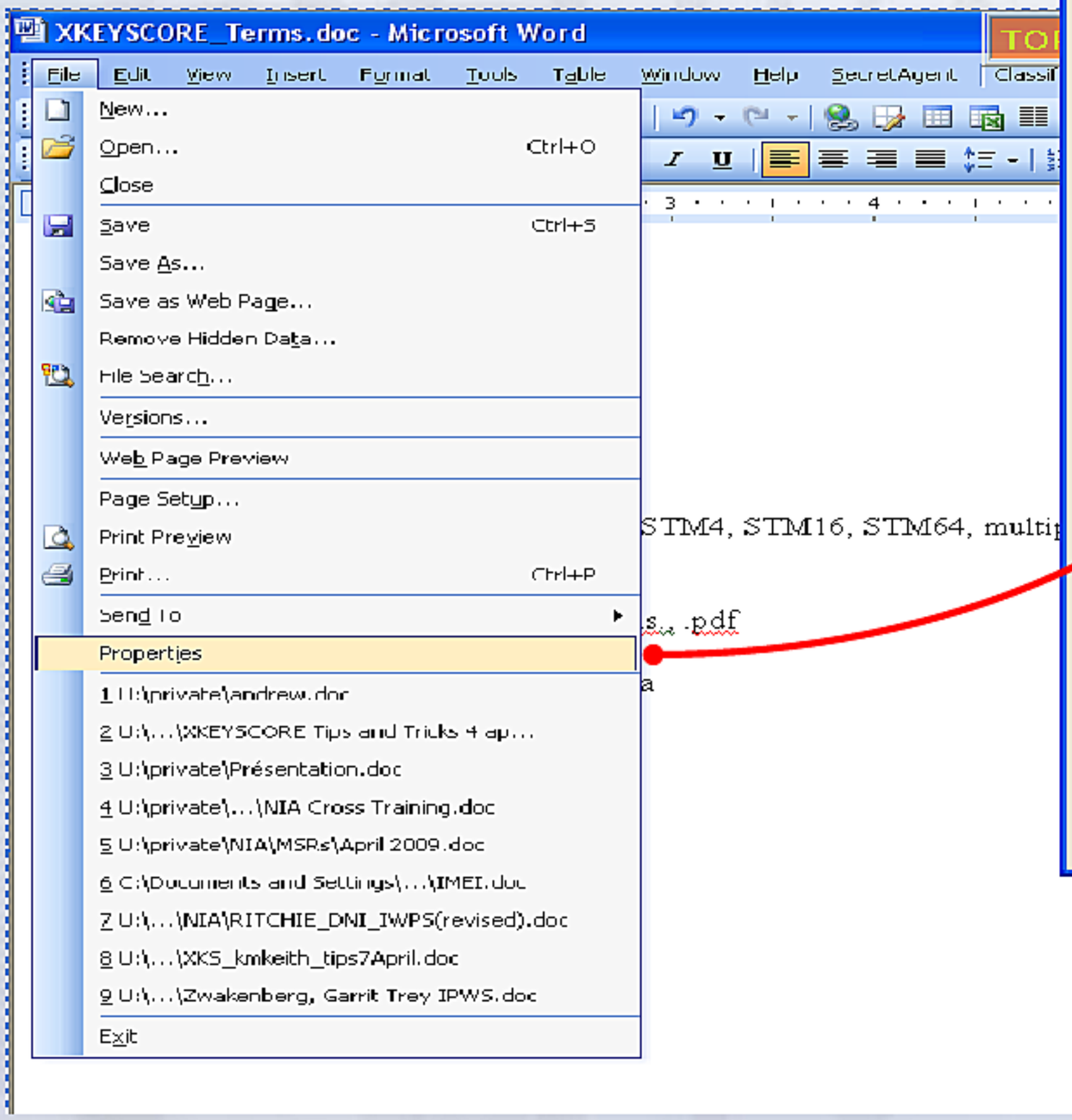
- Why to Query on Document Metadata
- How to Find Document Metadata
  - e.g. File - > Properties
  - Google
- How to Create Queries in XKS
  - XKEYSCORE Document Metadata and PDF Metadata

Document Metadata Analysis

- **What?**: Use *non-traditional* selectors to find and track targets sending/receiving documents of interest
- **How?** It targets documents by Author, Organization, or embedded images (logos)
- **Why?** We don't always know WHO is sending the documents, but they are "guilty-by-association" if they send/receive the document. So, who are THEY?

Finding Document Metadata

- We find "Document Metadata" in File Properties



Document Metadata Analysis

- How do you find document metadata?
  - Passive Collection: Collected Documents already contain data
  - Active Collection: CNE "Categorized Collection" from TUNINGFORK Data or Pinwale Queries on "US-3101"
  - Open Source: Google Hacking

Finding Document Metadata

KEYSCORE

ID	To	From	C(	B(	Date	Subject	Size(K)	Type
48	[redacted]	"DESTOCKPRO" [redacted]			5/15/2008 6:31:35 PM	ARRIVAGE G STAR DOLCE&GABBANA DIESEL	16	text/html
49	[redacted]	"Target Center" [redacted]			5/16/2037 9:36:13 PM	Confirmation: Target Card	2	text/html
51		[redacted]			5/15/2008 11:14:32 PM		1	text/html
41	[redacted]	"khalid" [redacted]			5/15/2008 3:05:25 PM	Las Villas de Dubai	3898	application/octet-stre
32	[redacted]	[redacted]			5/15/2008 11:15:21 PM	skriv ut	452	application/msword
							50	application/msword
51		[redacted]			5/15/2008 11:14:32 PM		452	application/msword
							50	application/msword

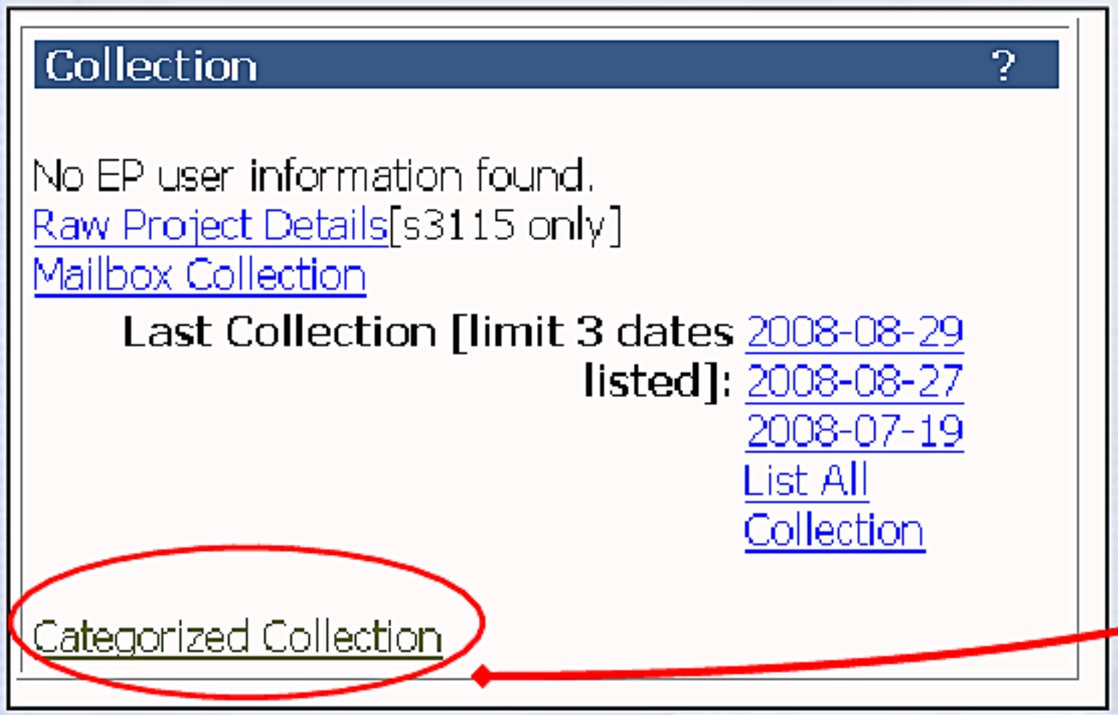
Display original | Raw | SMTP header | **Properties** | Control | C2C | Trailer | Collected Doc | Search Kwd | Dictionary Kwd | SRI

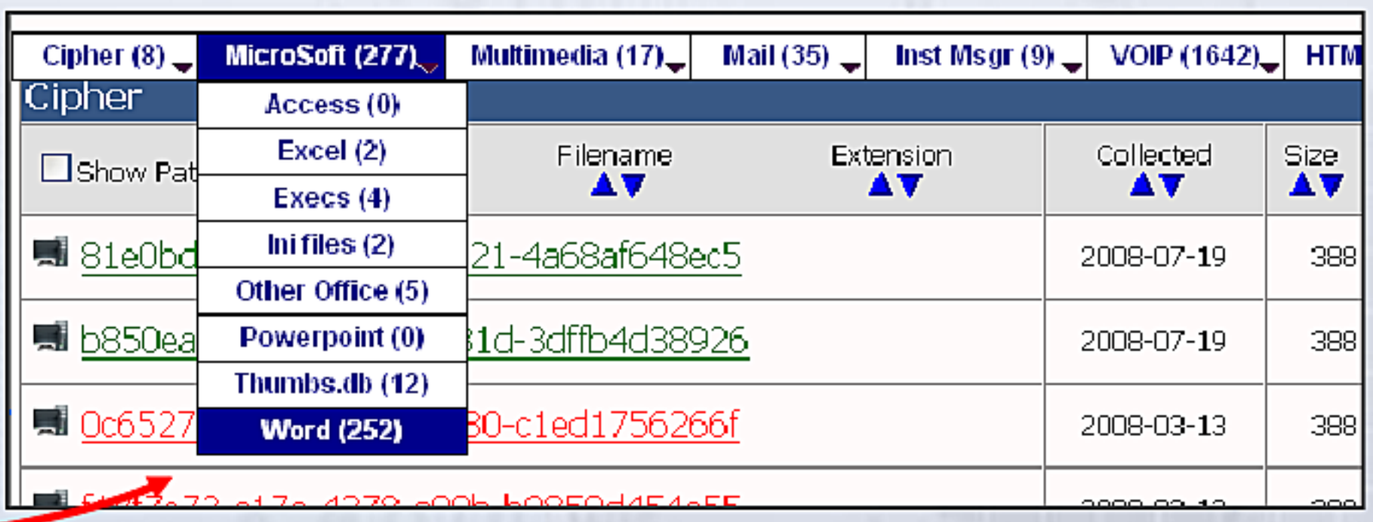
	3898	application/octet-strear	bas
	452	application/msword	bas
	50	application/msword	bas

Document Properties	
Category	
Company	
HiddenSlideCount	0
LineCount	29
LinksUpToDate	False
Manager	
MMClipCount	29
NoteCount	29
ParagraphCount	8
PresentationTarget	
ScaleCrop	False
SlideCount	8
AppName	Microsoft Word 10.0
Author	GoGo
CharacterCount	3582
Comments	
DateCreated	5/12/2008 3:13:00 AM
SecurityLevel	none
Keywords	
LastAuthor	[redacted]

Finding Document Metadata

- Active Collection: CNE "Categorized Collection" from TUNINGFORK Data





To find Document Metadata in TUNINGFORK, you must view each Document in Categorized Collection (manual intensive)

Finding Document Metadata

- Using XKEYSCORE to query on CNE data

Fields ▾ Advanced Features ▾ Show Hidden Search Fields Clear Search Values Reload Last Search Values

Search: Document Metadata

Input Source: FOXACID* or FOXBASE*

Selected implant exfils from active collection (xks-cne.corp.nsa.ic.gov:xs_web_db)

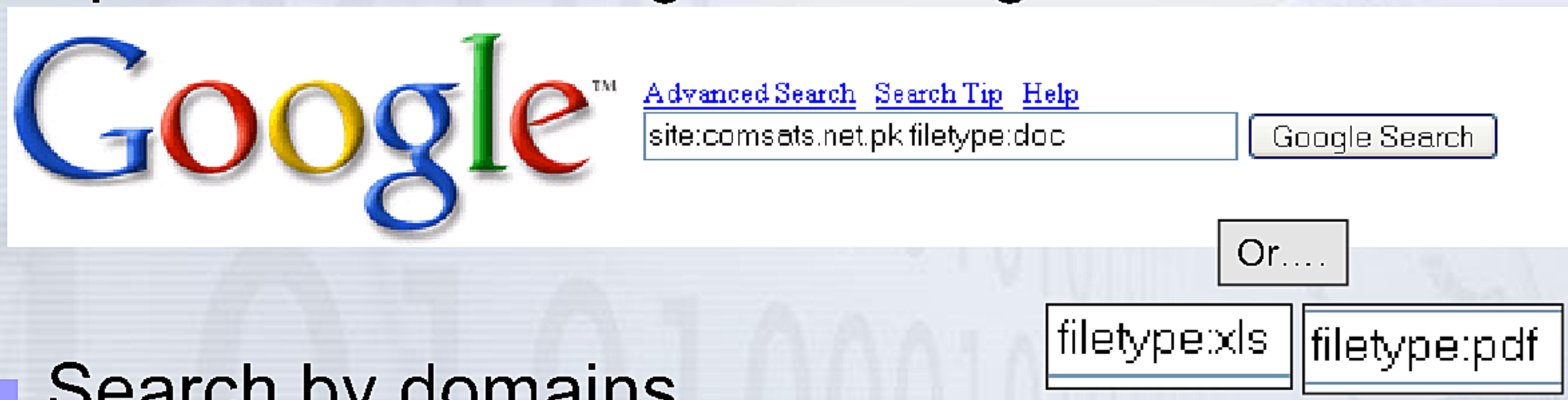
This query in XKS

Produced these results

Filename	Extension	Author	Title	Input Source
\System Admin CV.doc	doc	Authorised User		[redacted]E4
C:\Downloads\Physical_Layer_in_RPR_0402.pdf	pdf			[redacted]ER12
C:\Documents and Settings\Guest\Desktop\05070807_excelbook.pdf	pdf	Center For Excelence	(Microsoft Word - \307\341\343\336\3	
C:\Documents and Settings\user\Desktop\desktop icons\yementrack (	doc	user		[redacted]ER12
C:\Downloads\ns-3-overview.pdf	pdf			[redacted]ER12
C:\Documents and Settings\user\Desktop\desktop icons\servers_exp	doc	results	الأخ/ مدير عام المشتريات و المخازن المحترم	[redacted]ER12

Finding Document Metadata

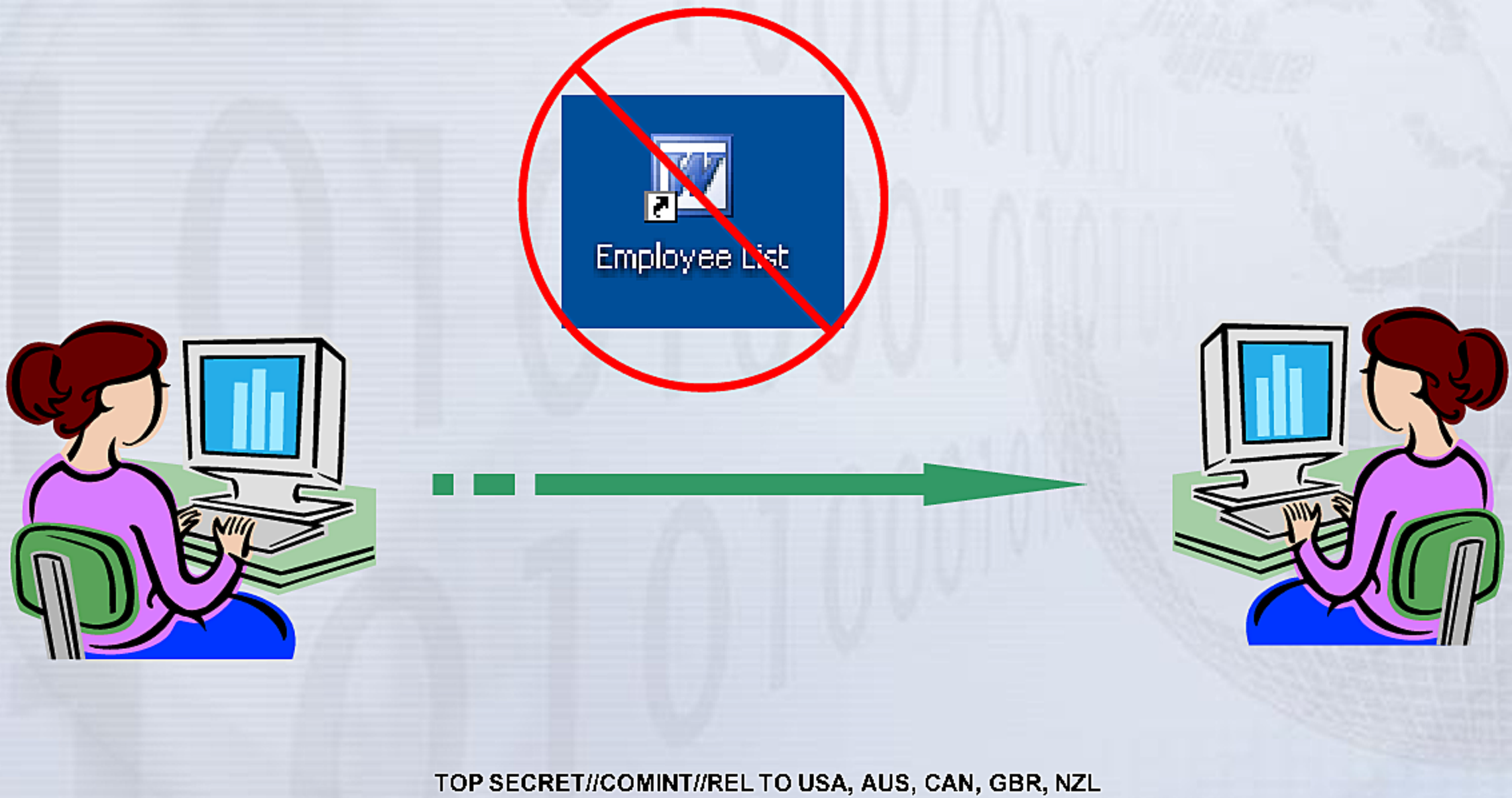
- Open Source: Google Hacking



- Search by domains
  - “site:comsats.net.pk”
- Search by file types
  - “filetype:pdf” or “filetype:doc”

Document Metadata Analysis

How to find Document Metadata when you have NEVER collected a document



Document Metadata Analysis

Take Client's (Active User) IP address and query on it in XKEYSCORE

Active User:

[redacted]@yahoo.com

Search: Document Metadata

Extension: ppt or doc or pdf or xls

IP Address: 89.[redacted] Either

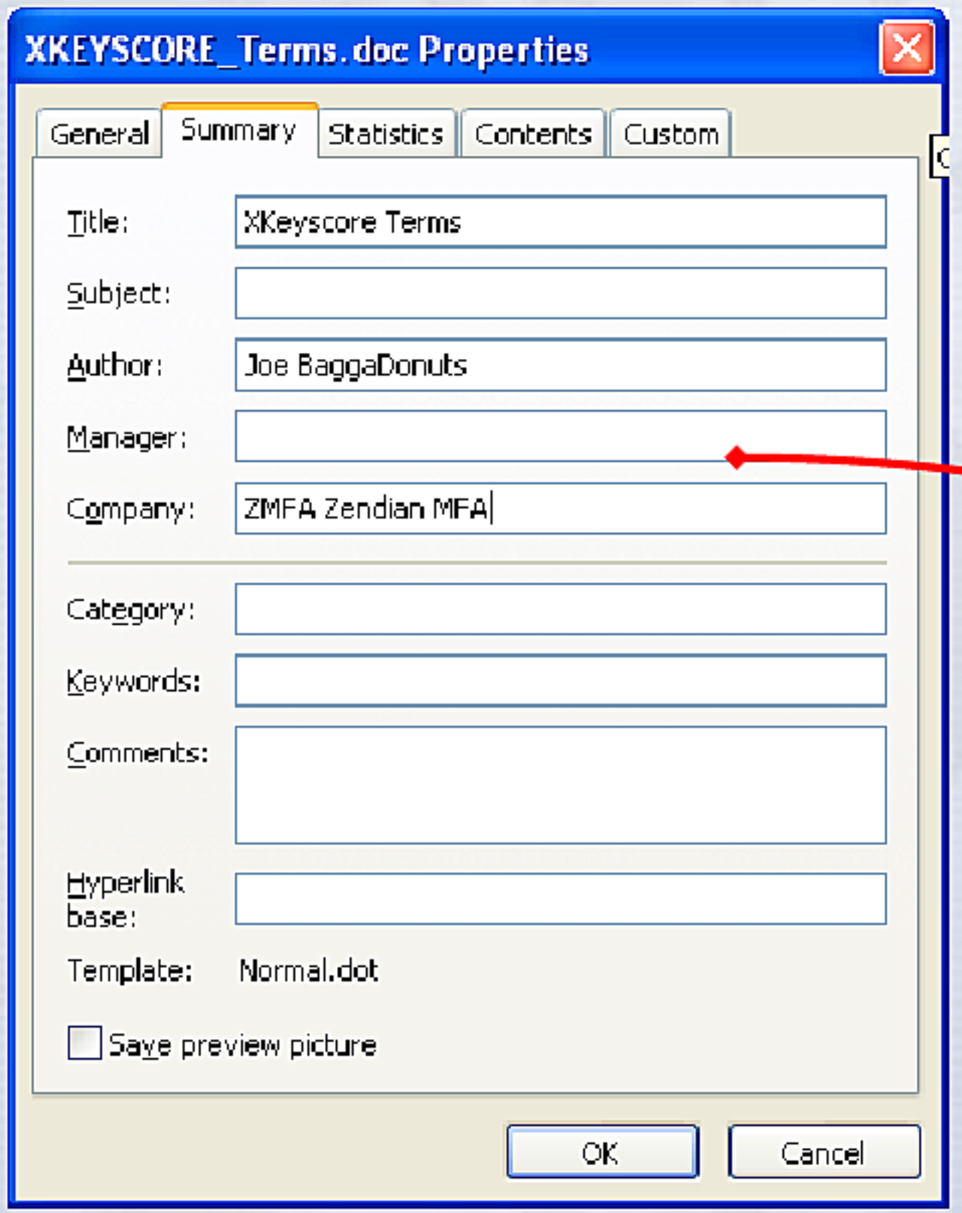
ACTIVE_USER	ACTIVE_USER_IP
[redacted] <yahoo>	89.[redacted]
[redacted] <yahoo>	89.[redacted]

Targeting Document Metadata

- Use XKEYSCORE to Find Who Else is sending the files?

Document Metadata Analysis

Take “File Properties” information and fill-in query



Search: Document Metadata

Document Type:	
Encrypted?:	
Corrupted?:	
Filename:	
Extension:	
Subject:	
Creation Time:	
Last Modified Time:	
Unique ID [fulltext]:	
Author:	Joe BaggaDonuts
Last Author:	
Organization:	ZMFA Zendian MFA
Title:	
Language:	
Comment [fulltext]:	
File/Embedded Image Hash [fulltext]:	
Metadata Name:	
Metadata Value [fulltext]:	

Document Metadata Analysis

Sample Query

Sample Query:

Organization = PTCL

To/From Country = Pakistan

Search: Document Metadata

Organization:	PTCL	
Title:		
Language:		
Comment [fulltext]:		
File/Embedded Image Hash [fulltext]:		
Metadata Name:		
Metadata Value [fulltext]:		
IP Address:		From
IP Address:		To
Port:		From
Port:		To
Country:	PK	Either

Document Metadata Analysis

Sample Query (Results)

Previous Slide produces these results

Filename	Organization
Instructions to Kunar province bidders community midwifery.doc	PTCL
Instructions to Kunar province bidders community midwifery.doc	PTCL

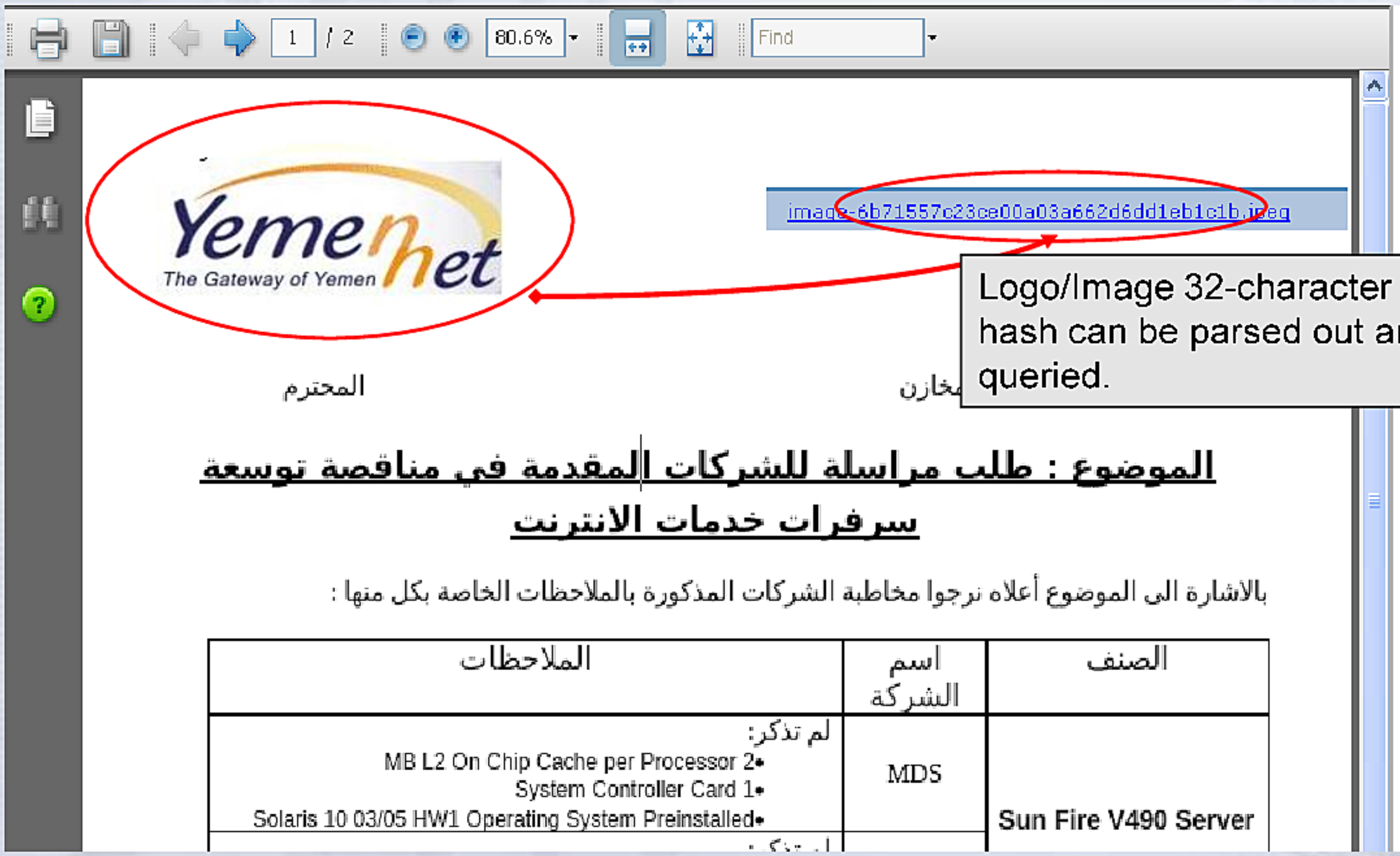
Embedded Images

- Turn a logo into a selector

= SIGINT VALUE

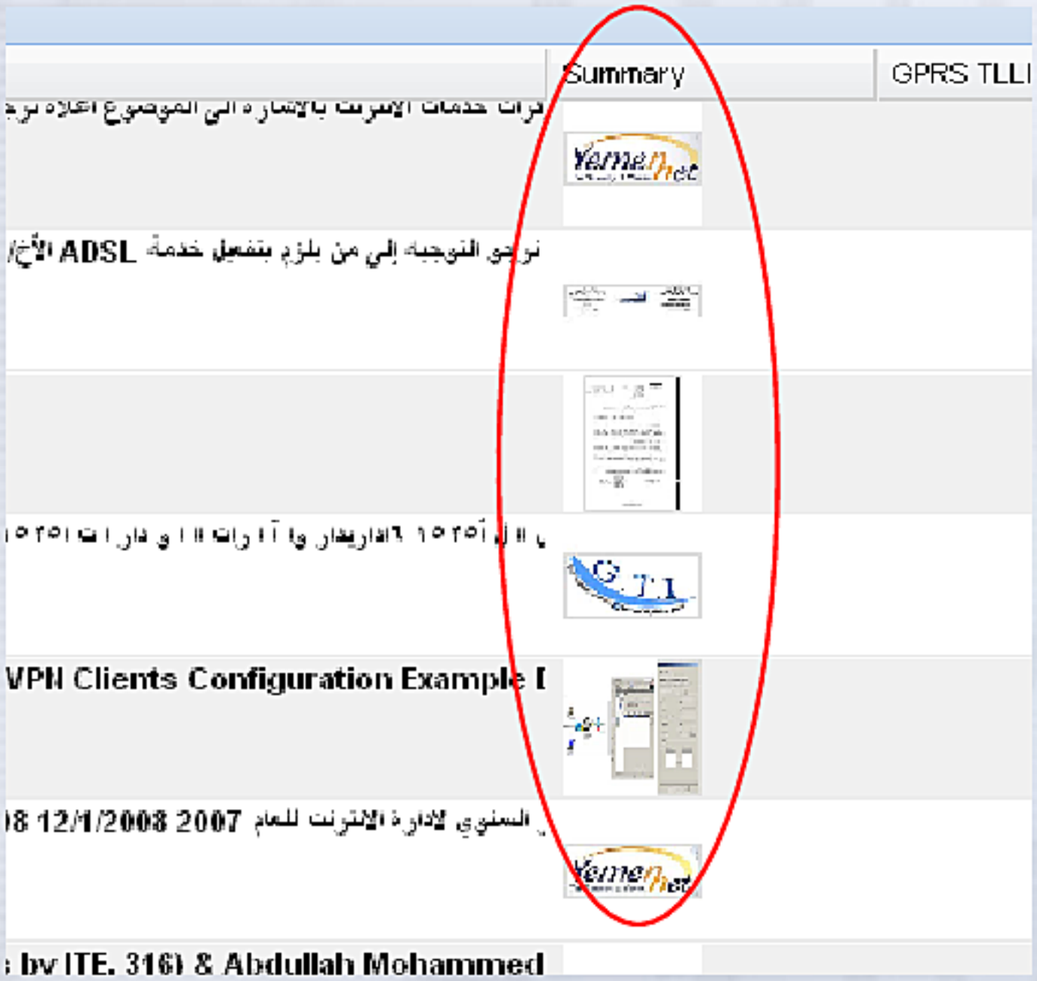
Embedded Images

- XKEYSCORE parses out logos from within documents (PDFs, DOCs, Outlook Emails, etc) embedded as images



Embedded Images

Files often contain embedded images, such as company logos.



Step 1: Identify if a document HAS an image in it

Embedded Images

Datetime	Case Notation	From IP
2009-03-26 15:54:50	YM.PGQXXXABDDTC	

Session | Header (3) | Attachments (6) | Meta (3)

Formatter: AUTO | Send to: Download Session | Mode: Full Session | Options

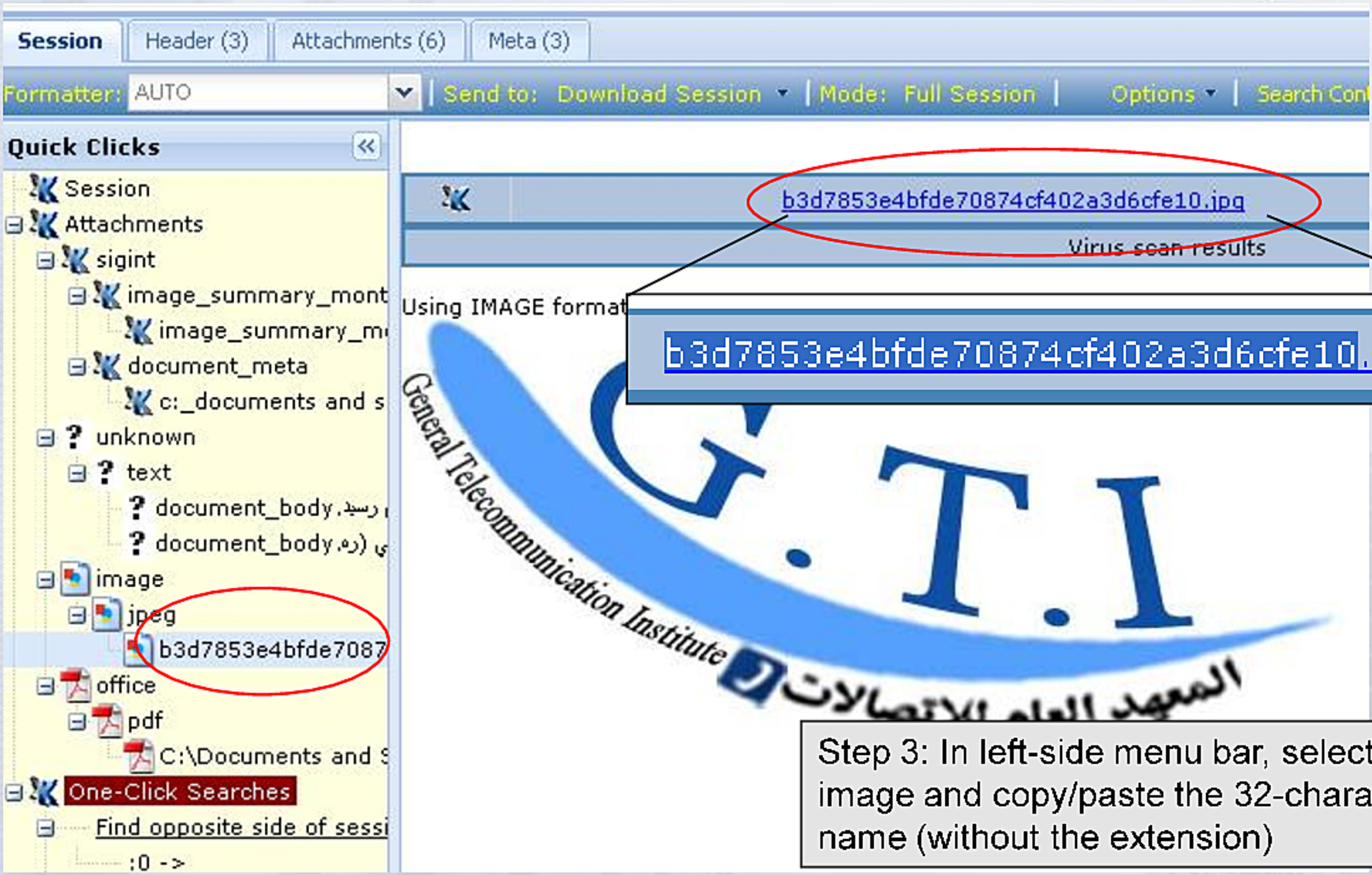
Quick Clicks

- Session
- Attachments
  - sigint
    - image_summary_mont
      - image_summary_m
    - document_meta
      - c:_documents and s
  - unknown
    - text
      - document_body.رسط
      - document_body.ه) ي
  - image
    - jpeg
      - b3d7853e4bfde7087
  - office
    - pdf
      - C:\Documents and S
- One-Click Searches
  - Find opposite side of sessi
    - :0 ->
      - :0
  - Find More Docs with Same
    - 635ed0657cfe25b7790f
    - b3d7853e4bfde70874cf

Retrieving Attachment...

Step 2: Open Document and click on “Full Session

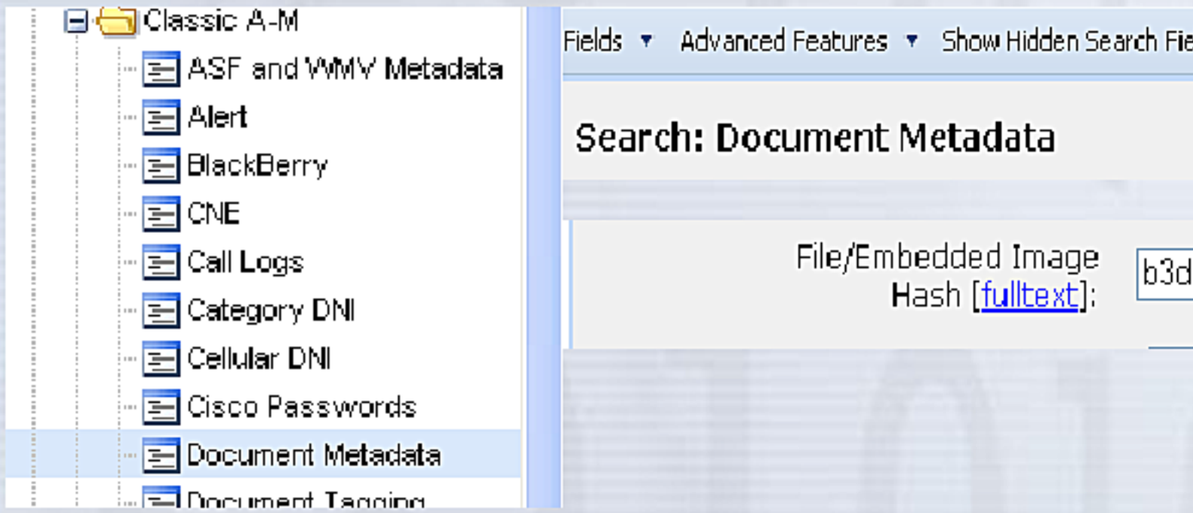
Embedded Images



Step 3: In left-side menu bar, select an image and copy/paste the 32-character name (without the extension)

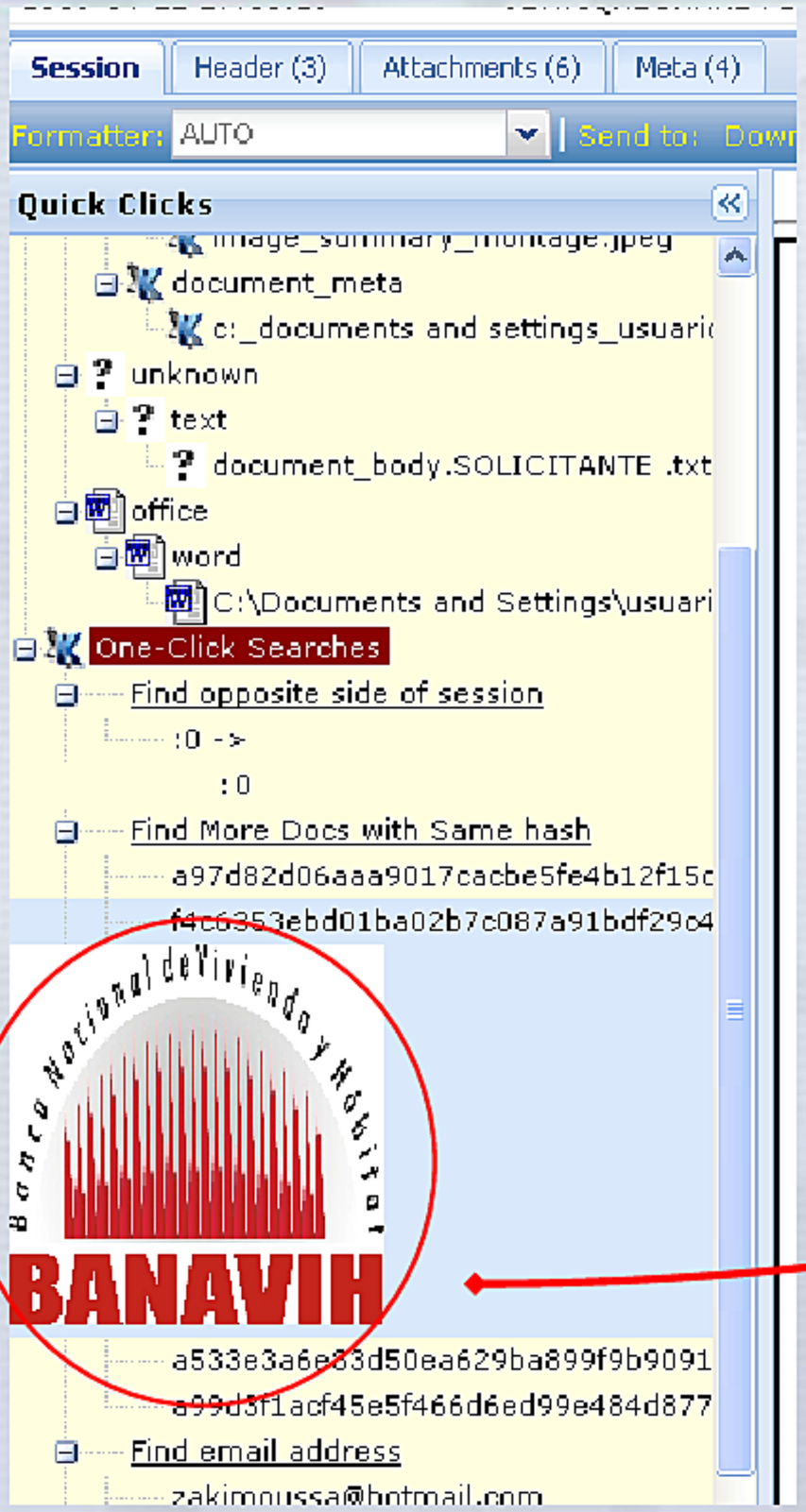
Embedded Images

Step 4: Paste the 32-character name into the "File/Embedded Image Hash" Field in the Document Metadata query

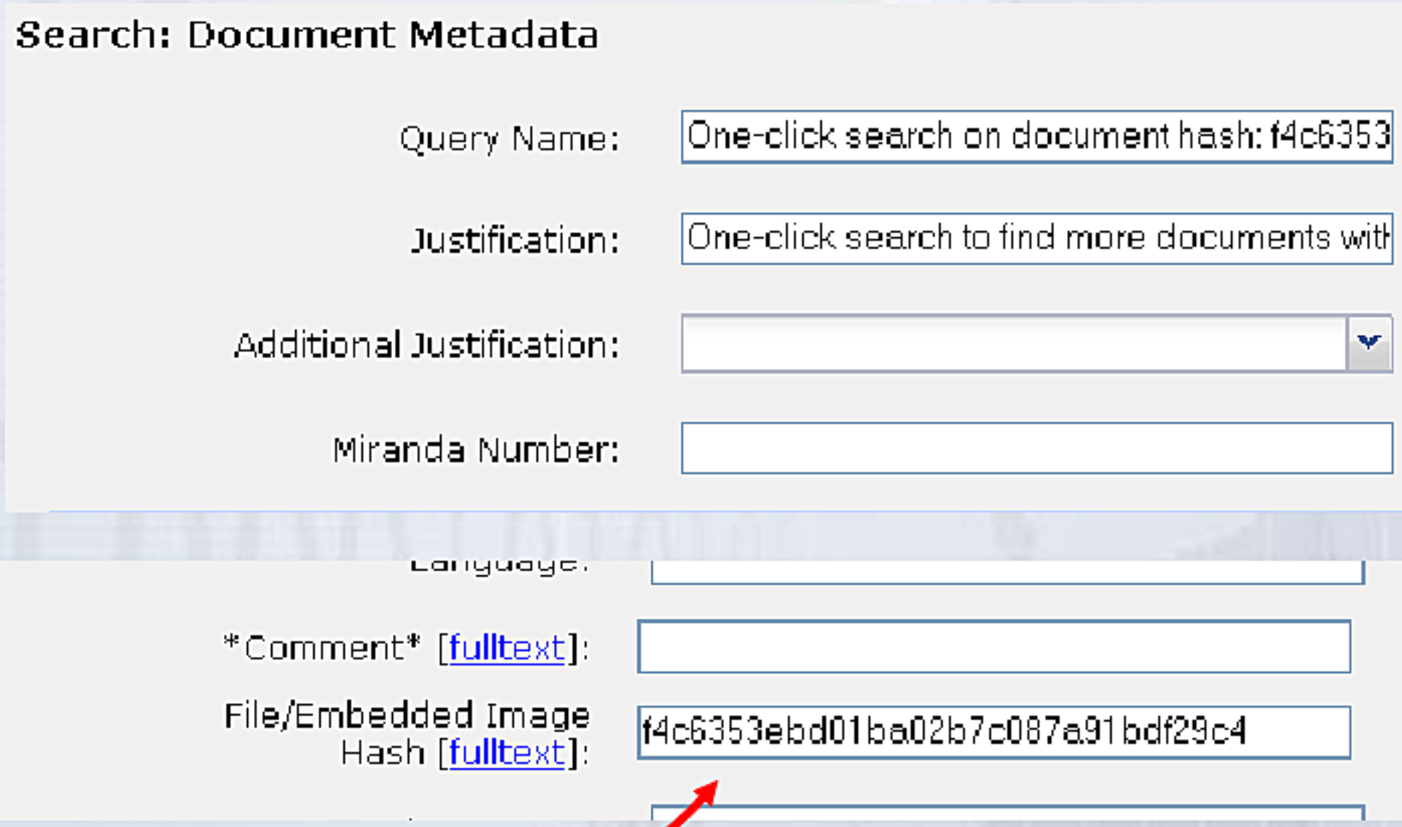


Step 5: Select all of your good collection sites + SUBMIT!

Embedded Images

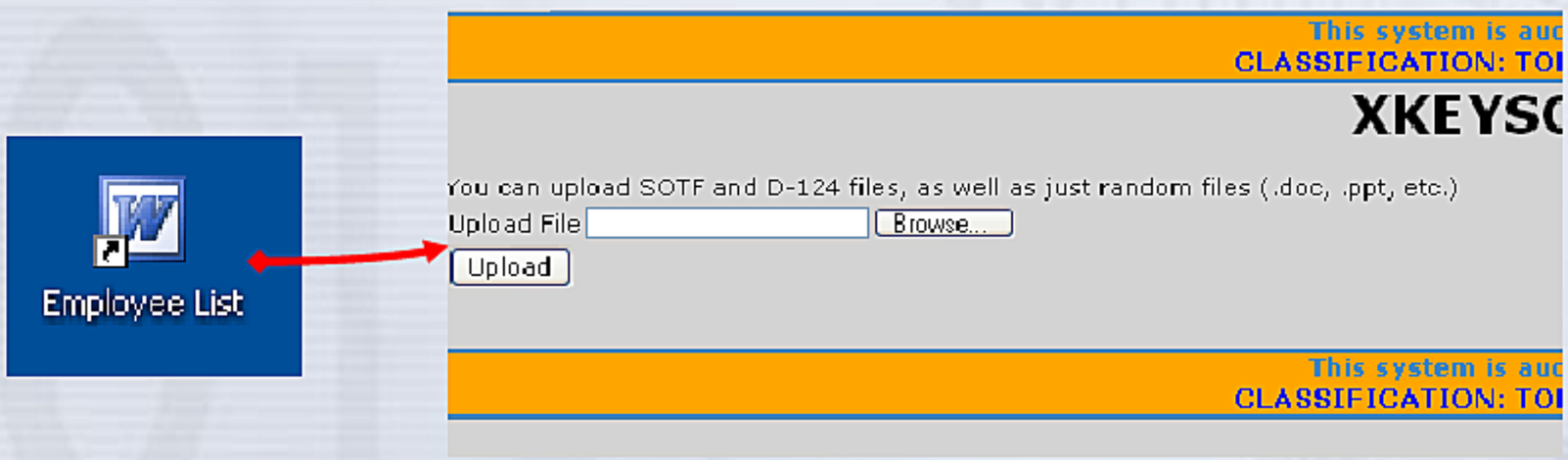


Or… You can one-click query to create a new query



Embedded Images

- Stand-alone files can be uploaded into XKS and images parsed out
  - Useful for TAO collection that didn't get into XKS (non United Rake)
  - https://xks-central.corp.nsa.ic.gov/general/view_file.php



- To task the hex values for images in CADENCE or Query in PINWALE, contact The Xtreme Target Pursuit Team **Tom Chaney S2I7** and **Sandra Farrier S3114**

Embedded Images

- Questions on any of these tools or techniques, contact:
  - [redacted]
  - [redacted]