Forenzika digitalnih dokumenata

Kristian Skračić
Predrag Pale

Što je forenzika?

- pronalaženje **tragova**
- utvrđivanje **slijeda** nastajanja tragova
- pridruživanje tragova **vremenu, mjestu i osobi**

- u svrhu
  - otkrivanja **ne-očiglednog**
    - onog što nije vidljivo **na prvi pogled**
    - onog što nije vidljivo **laicima**
  - pomaganja u **pronalaženju izgubljenog**
  - **pomaganja u istrazi** bilo koje vrste

Forenzika digitalnih dokumenata

- Digitalni dokument
  - je bilo kakva datoteka (eng. file)
  - koja sadrži informaciju
    - program, tekst, formatirani tekst, audio, video, slike,...
- Dokumenti su
  - osnovna apstrakcija/koncept za:
    - pohranu
    - manipuliranje
    - razmjenu
  - informacija na i između računala, programa, ljudi

- Stoga su dokumenti od velike važnosti
  - u prikupljanju digitalnih tragova

Što se želi postići

- Postupak forenzičke analize digitalnih dokumenata ima za cilj:
  1. prepoznati **vrstu sadržaja** datoteke
     - program, baza podataka
     - tekst, slika, audio, video, ….
  2. u datoteci pronaći i izvaditi **metapodatke**
     - koji opisuju datoteku
       - autora, vrijeme nastanka, mjesto ….
  3. u datoteci pronaći ostale "**nevidljive**" informacije
     - **obrisani** sadržaj (npr. za vrijeme uređivanja teksta)
     - **privremeni** sadržaj (npr. indeksi, tablice, …)
     - **ostatke** prijašnjih datoteka (koje su zauzimale isti prostor na disku)

IDENTIFIKACIJA SADRŽAJA

Identifikacija sadržaja

- Cilj:
  - prepoznati **vrstu sadržaja** datoteke

- Dvije su osnovne **metode**:
  - promatranje **ekstenzije** imena datoteke
    - .exe, .doc, .xls, .pdf, .jpg, .wmv …
  - **analizom strukture** sadržaja datoteke
    - **kakvi** se podaci nalaze na **specifičnim mjestima** u datoteci

Metoda 1: Ekstenzija imena datoteke

- metoda:
  - tekstualnu **oznaku**
    koja se nalazi iza posljednje točke u nazivu datoteke
    - .exe, .doc, .xls, .pdf, .jpg, .wmv …
  - treba **usporediti** s popisom poznatih ekstenzija

en.wikipedia.org/wiki/List_of_filename_extensions_(alphabetical)
http://filext.com

- čemu služi ekstenzija
  - osim **ljudima**,
    služi brojnim **programima** da odrede kako će postupati sa sadržajem neke datoteke
- kako nastaju ekstenzije
  - **autori** neke nove vrste sadržaja **odaberu** ekstenziju

- prednost metode:
  - **brzo i jednostavno** korištenje
  - ne treba otvarati sadržaj datoteke
- problem
  - ništa **ne sprečava** čovjeka ili program
  - **da promijene** ekstenziju datoteke **iako nisu promijenili** sadržaj
    - npr. .doc u .xls ili .exe u .pdf ili .jpg u .bin i sl.
  - stoga se **promjena ekstenzije** datoteke koristi kao **prvi korak u prikrivanju** pravog sadržaja datoteke
  - zato treba **prepoznati vrstu sadržaja** datoteke

Metoda 2: Analiza sadržaja datoteke

- metoda:
  - **otvara se** datoteka
  - **analizira se struktura** sadržaja i pojedina ključna mjesta
  - sadržaj tih mjesta se **uspoređuje s popisom poznatih** podataka
- ključna struktura je sam **početak datoteke**
  - gdje se nalazi **identifikacija vrste** sadržaja
  - tzv. "**magic number**"
- Prednost metode
  - bitno **pouzdanije** od ekstenzija
  - manipuliranje magic numbera je rijetko,
    jer onda normalni programi neće znati rukovati sadržajem datoteke
- Nedostatak metode
  - potrebno je **otvarati datoteku** i dohvaćati magic number
  - **potreban je popis** različitih magic number
  - **strukture** datoteka **ne moraju biti** javno **poznate**

https://en.wikipedia.org/wiki/List_of_file_signatures
http://www.asecuritysite.com/forensics/magic
http://www.garykessler.net/library/file_sigs.html
http://file-extension.net/seeker/

Primjer: PDF datoteka

```
00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F
25 50 44 46 2D 31 2E 34 0A 25 AA AB AC AD 0A 34   %PDF-1.4.%ª«¬..4
20 30 20 6F 62 6A 0A 3C 3C 0A 2F 54 69 74 6C 65    0 obj.<<./Title
```

- **Svaka** PDF datoteka **započinje** sekvencom (magic number) *"%PDF"*
  - tj. *0x2550 4446* (0x25 0x50 0x44 0x46)
- EXE datoteke započinju sekvencom ***0x4D5A***
- *itd.*

```
00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F
4D 5A 90 00 02 00 00 00 04 00 0F 00 FF FF 00 00   MZP.........ÿÿ..
B8 00 00 00 00 00 00 00 40 00 1A 00 00 00 00 00   ,.......@.......
```

Magic numbers

- Svaka datoteka započinje određenom sekvencom
  - i to nema čvrstu vezu s ekstenzijom datoteke
- Primjeri

Content type	Hex	ASCII
Windows exe	4D 5A	
Unix script	23 21	#!
MS office	D0 CF 11 E0	DOCFILE0
MS Office (OOXML) Open XML Format	50 4B 03 04 14 00 06 00	PK….
PDF	25 50 44 46	%PDF
GIF	47 49 46 38 39 61 47 49 46 38 37 61	GIF89a GIF87a
PNG	89 50 4E 47 0D 0A 1A 0A	\211PNG\r\n\032\n
JPEG	F8 D8	
Java binary	CA FE BA BE	
Java compressed	CA FE D0 0D	

- en.wikipedia.org/wiki/Magic_number_(programming)
- en.wikipedia.org/wiki/List_of_file_signatures
- http://www.garykessler.net/library/file_sigs.html
- http://file-extension.net/seeker/

Alati za interpretaciju

- Unix/Linux sustavi
  - naredba ***file***
    koja putem ključnih struktura može otkriti vrstu sadržaja
  - koristi popis ključnih struktura (engl. *magic number*) koji se nalazi na:
    - /etc/magic
    - /usr/share/misc/magic
  - Primjer zapisa ključne strukture:

    0 beshort 0xffd8 JPEG image data

- Windows
  - TrID http://mark0.net/soft-trid-e.html

```
$ file program
program: ELF 32-bit LSB executable, Intel 80386, version
1 (SYSV), dynamically linked (uses shared libs), stripped
```

Pregled sadržaja

- Tijekom forenzičke obrade, često koristimo specijalizirane alate zbog:
  - brzine
  - točnosti i dosljednosti
    - samog postupka
- No, niti jedan alat nije savršen
  - pa je u nekim slučajevima potrebno
  - ručno analizirati digitalne dokumente
- U tu svrhu se obično koriste **hexeditor** alati

Primjer hex editor alata:

File Edit Search View Analysis Extras Window ?

16 | ANSI | hex

forenzika_digitalnih_dokumenata.pptx

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	
00000000	50	4B	03	04	14	00	06	00	08	00	00	00	21	00	EE	C0	PK.....!.îÀ
00000010	2C	39	4D	02	00	00	11	12	00	00	13	00	08	02	5B	43	,9M.....[C
00000020	6F	6E	74	65	6E	74	5F	54	79	70	65	73	5D	2E	78	6D	ontent_Types].xm
00000030	6C	20	A2	04	02	28	A0	00	02	00	00	00	00	00	00	00	l ¢..(.....
00000040	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000050	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000060	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000070	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000080	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000090	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000000A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000000B0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000000C0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000000D0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000000E0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000000F0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000100	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000110	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000120	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000130	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000140	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000150	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000160	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000170	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000180	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
00000190	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000001A0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000001B0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000001C0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000001D0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....
000001E0	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	.....

Prikaz sadržaja u “*hex*” zapisu

Prikaz sadržaja u ASCII zapisu

- Ispisuju se samo printabilni znakovi
- Ostali se prikazuju u obliku točke

Utvrdite kakve su ovo datoteke

1

```
Offset(h)  00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00000000   50 4B 03 04 14 00 06 00 08 00 00 00 21 00 30 C9  PK.........!.0É
00000010   28 0C 72 01 00 00 A5 05 00 00 13 00 08 02 5B 43  (.r...Ą.......[C
00000020   6F 6E 74 65 6E 74 5F 54 79 70 65 73 5D 2E 78 6D  ontent_Types].xm
00000030   6C 20 A2 04 02 28 A0 00 02 00 00 00 00 00 00 00  l ˘..( .........
00000040   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
00000050   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
```

2

```
Offset(h)  00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00000000   89 50 4E 47 0D 0A 1A 0A 00 00 00 0D 49 48 44 52  ‰PNG........IHDR
00000010   00 00 01 18 00 00 01 18 08 02 00 00 00 08 EC 7E  ..............ě~
00000020   DB 00 00 00 06 74 52 4E 53 00 FE 00 01 00 FD 5B  Ű....tRNS.ţ...ý[
00000030   6C 0D 3B 00 00 05 C1 49 44 41 54 78 9C ED DD 41  l.;...ÁIDATxśíÝA
00000040   8E 23 37 10 00 41 CB F0 FF BF 3C 7B F0 99 07 EE  Ž#7..AËđ˙ż<{đ™.î
00000050   66 99 A4 1C 71 B5 D0 DD 96 26 41 60 0B 64 7F 7E  f™¤.qµĐÝ–&A`.d.~
```

3

```
Offset(h)  00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00000000   25 50 44 46 2D 31 2E 35 0D 0A 25 B5 B5 B5 B5 0D  %PDF-1.5..%µµµµ.
00000010   0A 31 20 30 20 6F 62 6A 0D 0A 3C 3C 2F 54 79 70  .1 0 obj..<</Typ
00000020   65 2F 43 61 74 61 6C 6F 67 2F 50 61 67 65 73 20  e/Catalog/Pages
00000030   32 20 30 20 52 2F 4C 61 6E 67 28 68 72 2D 48 52  2 0 R/Lang(hr-HR
00000040   29 20 2F 53 74 72 75 63 74 54 72 65 65 52 6F 6F  ) /StructTreeRoo
00000050   74 20 33 32 20 30 20 52 2F 4D 61 72 6B 49 6E 66  t 32 0 R/MarkInf
```

4

```
Offset(h)  00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00000000   4D 5A 90 00 03 00 00 00 04 00 00 00 FF FF 00 00  MZ..........˙˙..
00000010   B8 00 00 00 00 00 00 00 40 00 00 00 00 00 00 00  ¸.......@.......
00000020   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
00000030   00 00 00 00 00 00 00 00 00 00 00 00 80 00 00 00  ............€...
00000040   0E 1F BA 0E 00 B4 09 CD 21 B8 01 4C CD 21 54 68  ..ş..´.Í!¸.LÍ!Th
00000050   69 73 20 70 72 6F 67 72 61 6D 20 63 61 6E 6E 6F  is program canno
```

Tip sadržaja	Hex
EXE Windows	4D 5A
SH Unix script	23 21
MSO ffice	D0 CF 11 E0
MSO OOXML	50 4B 03 04 14 00 06 00
PDF	25 50 44 46
GIF	47 49 46 38 39 61 47 49 46 38 37 61
PNG	89 50 4E 47 0D 0A 1A 0A
JPEG	F8 D8
JAR Java binary	CA FE BA BE
JAR Java compress	CA FE D0 0D

Ali, što ako ...

- nam **ključne** strukture **nisu dostupne**
  - obrisane su radi prikrivanja tragova
  - uništene/oštećene su prilikom pribavljanja datoteke
  - ...
- Tada je potrebno **tražiti**
  druge **poznate značajke u** samim **podacima**
  - **metapodaci** (eng. *Metadata*)

EKSTRAKCIJA MEDATPODATAKA

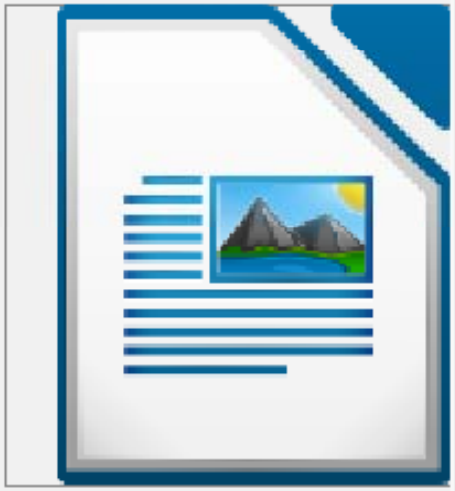
Metapodaci - podaci o podacima

- Mnogi dokumenti **uz sadržaj** pohranjuju **dodatne informacije** (metapodatke) koji dodatno **opisuju informacije** koje se nalaze u dokumentu

- Najčešće dodatno opisuju ili proširuju informaciju koju daje sam sadržaj datoteke
  - Primjeri:
    - timestamp, (su)autori, verzija, lozinke, lokacija, ...

Metapodaci - primjer

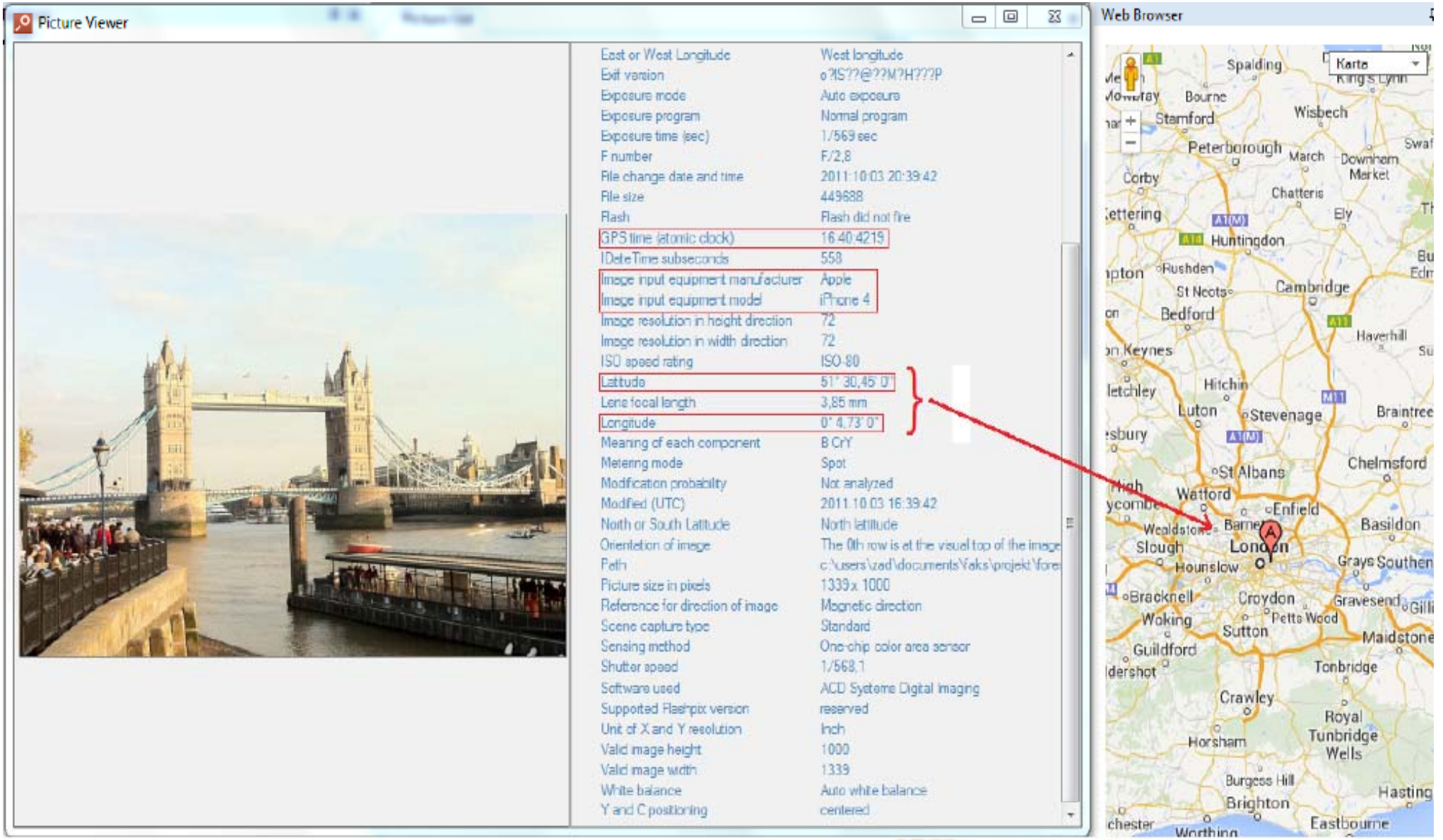
Picture Viewer



Created (UTC)	2014.10.09 14:13:17
Document text	SVEUČILIŠTE U ZAGREBUFAKULTET ELEK
File size	1262080
Last date the document was printed	2007.02.23 10:47:00
Last time the document was saved	2007.02.23 11:50:00
Modified (UTC)	2014.05.21 20:37:13
Number of pages	0
Number of words	0
Path	c:\users\zad\documents\faks\projekt\forenzik
SingleDocument	
The author of the document	Sinisa Tomic
The creation time of the document	2007.02.15 15:51:00
The last author of the document	Mario Cifrek
The minutes of editing the document	61
Title	Sveučilište u Zagrebu

Seminar iz 2014. napravljen po predlošku još ima podatke o vremenu izrade i autorima predloška iz 2007.!

Metapodaci – lokacija u slici



Oprez!

- Metapodaci nisu uvijek dostupni
  - Neki **stariji** datotečni **sustavi ne podržavaju** metapodatke
  - Metapodaci **nisu nužno potrebni** za prikaz/obradu/pohranu sadržaja, pa ih **neki alati** i sustavi **brišu**

- Metapodaci se mogu ***obrisati i izmijeniti***
  - bez da se promijeni sadržaj!
  - Ovisno o istrazi, već i to što nema metapodataka može biti indicija!

Primjer alata: Hachoir metadata

$ hachoir-metadata *pacte_des_gnous.avi*

Common:
- Duration: 4 min 25 sec
- Comment: Has audio/video index (248.9 KB)
- MIME type: video/x-msvideo
- Endian: Little endian

Video stream:
- Image width: 600
- Image height: 480
- Bits/pixel: 24
- Compression: DivX v4 (fourcc:"divx")
- Frame rate: 30.0

Audio stream:
- Channel: stereo
- Sample rate: 22.1 KHz
- Compression: MPEG Layer 3

ANALIZA DIGITALNIH DOKUMENATA

Dokumenti

- “dokumenti” = generički pojam
- no u ovom kontekstu podrazumijeva
  - datoteku koja sadrži:
    - tekst, slike
    - te informacije za prikaz sadržaja (engl. *rendering*)
  - na primjer:
    - PDF, DOC/DOCX, XLS/XLSX, ODF ...

Motivacija za analizu dokumenata

- Dokumenti su vrlo važni u forenzičkoj istrazi
  - **obilje** informacija o:
    - autorstvu, revizijama,
    - vremenske oznake,
    - operacijski sustavi na kojima se koristio ...
- Primjer iz stvarnog svijeta:
  - Serijski **ubojica**
    poznat pod nazivom *"BTK Killer" (Bind, Torture, Kill)*,
    otkriven je upravo uz pomoć metapodataka u MS Word dokumentu
    - U veljači 2005. godine,
      BTK je poslao disketu medijima i policiji
      u kojoj preuzima odgovornost za svoje žrtve
    - Analizom metapodataka otkriveno je:
      - Adresa na kojoj je dokument stvoren (*Christ Lutheran Church*)
      - Ime osobe koja je posljednja mijenjala datoteku (*Dennis*)

https://precisioncomputerinvestigations.wordpress.com/2010/04/14/how-computer-forensics-solved-the-btk-killer-case/

Najčešći razlozi analize digitalnih dokumenata

- Dokazivanje autentičnosti dokumenta
  - autora
  - vremena nastanka/izmjene
  - …
- Pronalaženje podataka koji donose dodatne informacije u istragu
  - **tko** je sve radio na dokumentu
  - **koliko dugo** se radilo
  - koliko puta je **dorađivan**
  - **gdje** je dokument nastao
  - ali također i za
    - obrisane sadržaje iz (dijela) datoteke
    - sadržaji iz drugih datoteka koje su ranije koristile isti medij
- Je li **dokument** korišten za **dostavu zlođudne** aplikacije

Formati dokumenata

- ASCII – TXT
- OLE
- OOXML
- ODF
- PDF
- Images
- Audio
- Video

ASCII – TXT datoteka

- Sadrži “čisti” tekst
- Bez podataka o formatiranju
- Bez metapodataka

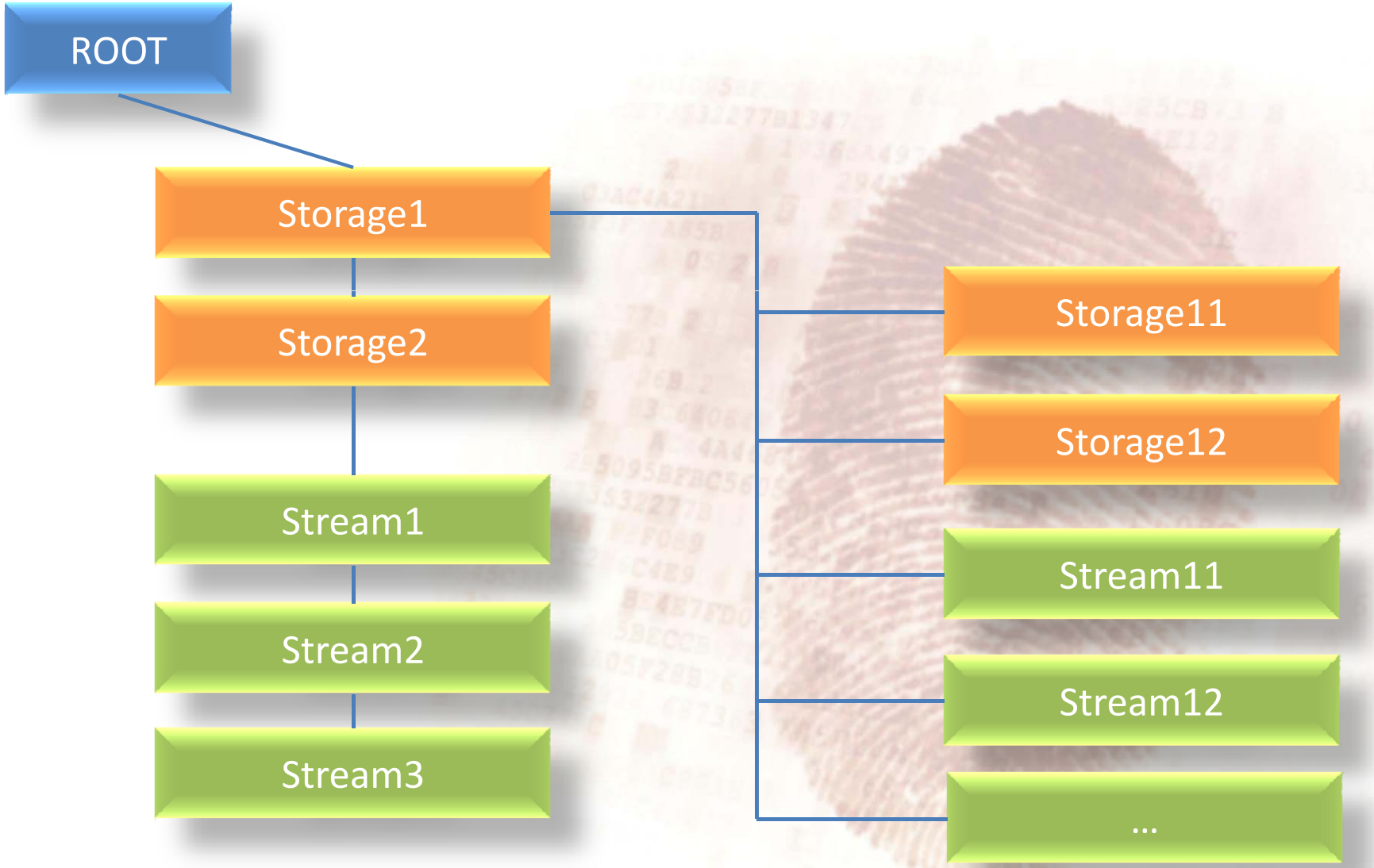
```
Offset(h)  00 01 02 03 04 05 06 07 08 09 0A 0B 0C 0D 0E 0F

00000000   54 68 69 73 20 69 73 20 61 20 73 69 6D 70 6C 65  This is a simple
00000010   20 66 69 6C 65 20 69 6E 20 41 53 43 49 49 20 66   file in ASCII f
00000020   6F 72 6D 61 74 2C 20 66 72 65 71 75 65 6E 74 6C  ormat, frequentl
00000030   79 20 63 61 6C 6C 65 64 20 84 54 58 54 93 2E 0D  y called „TXT“..
00000040   0A 44 6F 65 73 20 6E 6F 74 20 63 6F 6E 74 61 69  .Does not contai
00000050   6E 20 61 6E 79 20 74 65 78 74 20 66 6F 72 6D 61  n any text forma
00000060   74 74 69 6E 67 20 69 6E 66 6F 72 6D 61 74 69 6F  tting informatio
00000070   6E 2E 0D 0A 44 6F 65 73 20 6E 6F 74 20 63 6F 6E  n...Does not con
00000080   74 61 69 6E 20 61 6E 79 20 6D 65 74 61 64 61 74  tain any metadat
00000090   61 2E 0D 0A 0D 0A 4F 76 6F 20 6A 65 20 6F 62 69  a.....Ovo je obi
000000A0   E8 6E 61 20 64 61 74 6F 74 65 6B 61 20 75 20 41  čna datoteka u A
000000B0   53 43 49 49 20 66 6F 72 6D 61 74 75 2C 20 E8 65  SCII formatu, če
000000C0   73 74 6F 20 6E 61 7A 76 61 6E 61 20 84 54 58 54  sto nazvana „TXT
000000D0   93 2E 20 0D 0A 4E 65 20 73 61 64 72 9E 69 20 6E  “. ..Ne sadrži n
000000E0   69 6B 61 6B 76 6F 20 66 6F 72 6D 61 74 69 72 61  ikakvo formatira
000000F0   6E 6A 65 20 74 65 6B 73 74 61 2E 0D 0A 4E 65 20  nje teksta...Ne
00000100   73 61 64 72 9E 69 20 6D 65 74 61 70 6F 64 61 74  sadrži metapodat
00000110   6B 65 2E 0D 0A 0D 0A                             ke.....
```

OLE datoteke

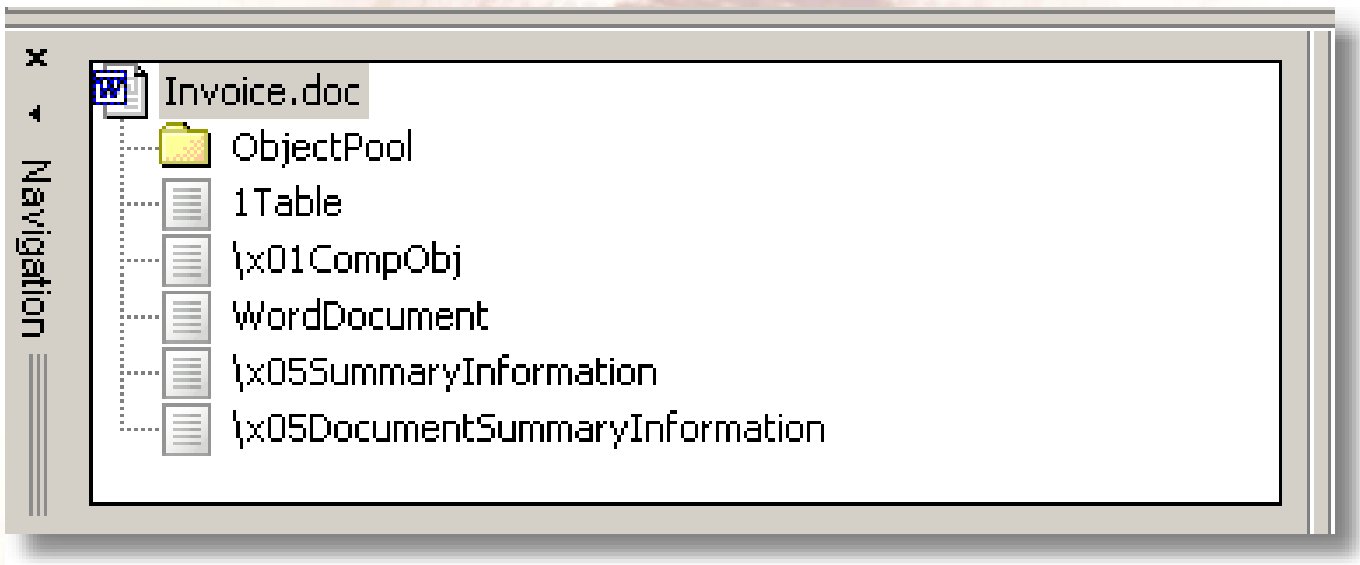
- *OLE Compound files*
  - Format koji koristi velik broj datoteka:
    - DOC, XLS (općenito MS Office do inačice 2003)
  - Mali, **prenosivi** datotečni sustav
  - Dva koncepta pohrane podataka:
    - **Storage** objects
    - **Stream** objects
  - Kao svi datotečni sustavi, sadrži
    - jedan **vršni** (root) direktorij,
    - te barem jedan ***stream*** objekt

OLE primjer


ROOT
Storage1
Storage2
Stream1
Stream2
Stream3
Storage11
Storage12
Stream11
Stream12
...

OLE – DOC primjer

Value (prefix)	Description
0x00	empty
0x01	storage
0x02	stream
0x03	lock bytes
0x04	property
0x05	root storage



OLE metapodatci

- Informacije o autorstvu
  - Korisničko ime osobe koja je dokument:
    - prva stvorila
    - posljednja uređivala
- Komentari
- Povijest revizija
- Količina vremena utrošena za obradu dokumenta
- Razne vremenske oznake (timestamp)

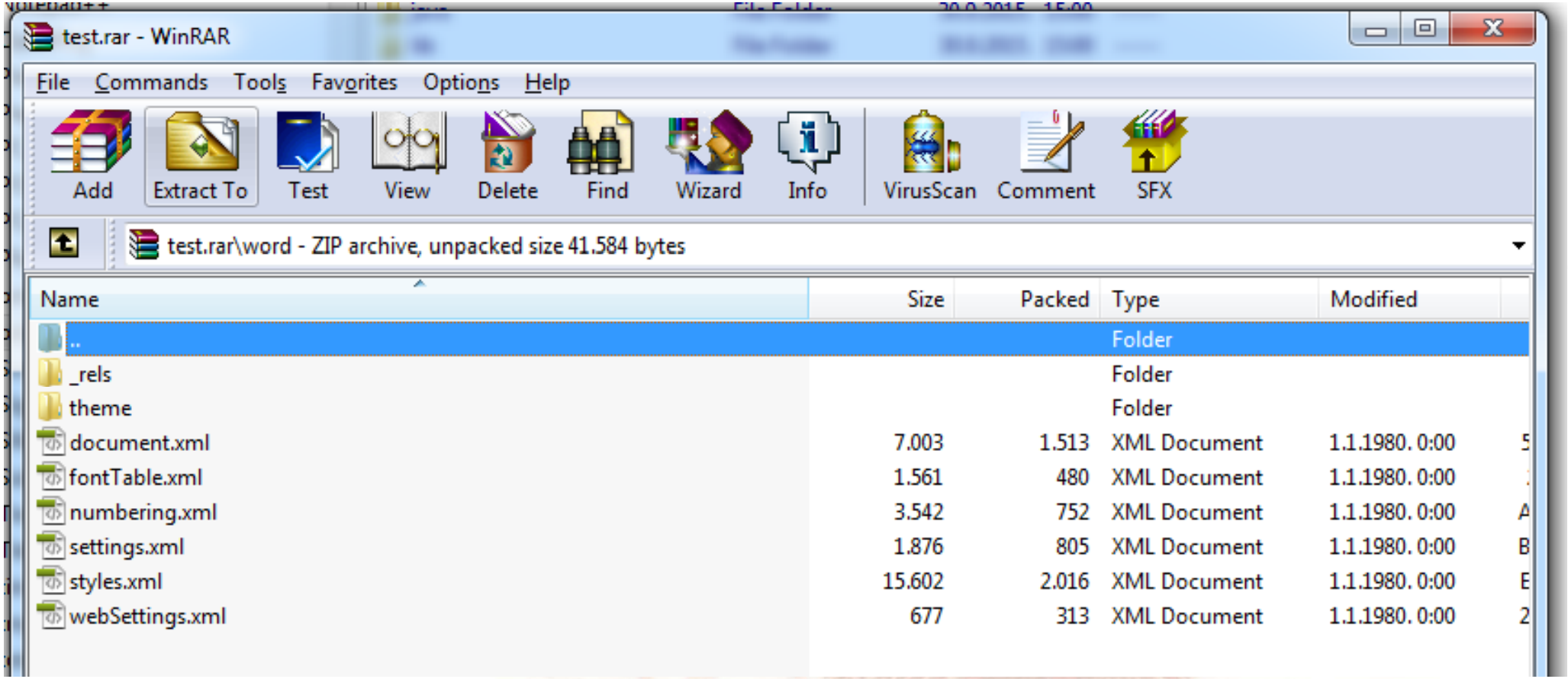
Primjer alata – Libforensics biblioteka

- Libforensics (olels.py)
  - Open-source biblioteka s raznim alatima koji **pomažu u** forenzičkoj **analizi**
    - **olels.py** – alat koji omogućuje analizu OLE datoteka
  - Drugi zanimljivi/korisni alati:
    - **datedecoder.py** – dekodira razne timestamp formate
    - **info2ls.p** – ispisuje sadržaj INFO2 (recycle bin) datoteka
    - **tdbcat.py** – ekstrahira thumbnail slike iz thumbs.db datoteka
    - **wmg.py** – ekstrahira metapodatke iz Microsoft Word dokumenata
    - …

Office Open XML – OOXML

- Jedan od trenutno **najpopularnijih** formata za pohranu **izmjenjivih** dokumenata
  - DOCX, XLSX, PPTX ...
  - **Izgled** dokumenta se **definira u trenutku** gledanja/ispisa
  - Sadržaj dokumenta se može mijenjati
- *Microsoftova zamjena za OLE*
- Kao i OLE,
  OOXML u sebi **sadrži niz dodatnih datoteka**
  koje se koriste za prikaz dokumenta
- Za razliku od OLE formata,
  OOXML **koristi ZIP arhivu** za pohranu sadržaja

DOCX Primjer - WinRAR



DOCX primjer: unzip

```
root@kali:~# unzip '/tmp/VirtualBox Dropped Files/2015-09-
17T12:38:54.407916000Z/forenzika_digitalnih_dokumenata.pptx'

Archive:  /tmp/VirtualBox Dropped Files/2015-09-
17T12:38:54.407916000Z/forenzika_digitalnih_dokumenata.pptx
  inflating: [Content_Types].xml
  inflating: _rels/.rels
  inflating: ppt/slides/_rels/slide13.xml.rels
  inflating: ppt/_rels/presentation.xml.rels
  inflating: ppt/presentation.xml
  inflating: ppt/slides/slide1.xml
  inflating: ppt/slideMasters/slideMaster1.xml
  inflating: ppt/slideLayouts/_rels/slideLayout10.xml.rels
  inflating: ppt/slideLayouts/slideLayout11.xml
  inflating: ppt/slideLayouts/_rels/slideLayout11.xml.rels
 extracting: ppt/media/image1.png
  inflating: ppt/theme/theme1.xml
 extracting: docProps/thumbnail.jpeg
 extracting: ppt/media/image4.gif
 extracting: ppt/media/image2.png
 extracting: ppt/media/image3.png
  inflating: ppt/tableStyles.xml
```

docProps/core.xml

- Sadrži glavni dio metapodataka

```
root@kali:~/docProps# cat core.xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<cp:coreProperties xmlns:cp="...">
        <dc:title>Forenzika digitalnih dokumenata</dc:title>
        <dc:creator>kiki</dc:creator>
        <cp:lastModifiedBy>kiki</cp:lastModifiedBy>
        <cp:revision>65</cp:revision>
        <dcterms:created xsi:type="dcterms:W3CDTF">
                2006-08-16T00:00:00Z
        </dcterms:created>
        <dcterms:modified xsi:type="dcterms:W3CDTF">
                2015-09-14T21:15:34Z
        </dcterms:modified>
</cp:coreProperties>
```

docProps/app.xml

- Dodatni (prošireni) metapodatci

```
…       <TotalTime>221</TotalTime>
        <Words>1095</Words>
        <Application>Microsoft Office PowerPoint</Application>
        <PresentationFormat>On-screen Show (4:3)</PresentationFormat>
        <Paragraphs>209</Paragraphs>
        <Slides>38</Slides>
        <Notes>0</Notes>
        <HiddenSlides>0</HiddenSlides>
        <MMClips>0</MMClips>
        <ScaleCrop>false</ScaleCrop>
        …
        <Company/>
        <LinksUpToDate>false</LinksUpToDate>
        <SharedDoc>false</SharedDoc>
        <HyperlinksChanged>false</HyperlinksChanged>
        <AppVersion>12.0000</AppVersion>
</Properties>
```

Napomena!

- Sve datoteke koje su **sastavni dio** dokumenta
  - slike, visio dokument, audio/video ...
  - I dalje **mogu sadržavati** metapodatke koje su **svojstvene za njihov tip** formata
    - na primjer:
      JPEG će i dalje imati svoje EXIF informacije unutar OOXML

ODF - OpenDocument Format

- Format sličan OOXML-u
- Također koristi ZIP arhivu
- Koristi se u OpenOffice alatu, te njegovim derivatima
  - Libre Office
- Podržava ga i
  - Microsoft Office 2010 supports ODF 1.1
  - Microsoft Office 2013 supports ODF 1.2
  - te 20-tak drugih “office” alata

Primjer ODF sadržaja

```
root@kali:~/Downloads# unzip Create_A_Template.odt
Archive:  Create_A_Template.odt
 extracting: mimetype
 extracting: Pictures/10000000000006900000028340F4295.jpg
 extracting: Pictures/10000000000001BE0000005774B9E5A7.gif
  inflating: content.xml
  inflating: layout-cache
  inflating: styles.xml
 extracting: meta.xml
  inflating: Thumbnails/thumbnail.png
  inflating: Configurations2/accelerator/current.xml
   creating: Configurations2/progressbar/
   creating: Configurations2/floater/
   creating: Configurations2/popupmenu/
   creating: Configurations2/menubar/
   creating: Configurations2/toolbar/
   creating: Configurations2/images/Bitmaps/
   creating: Configurations2/statusbar/
  inflating: settings.xml
  inflating: META-INF/manifest.xml
```

meta.xml

- Zanimljivost:
  - Sadrži točan “**build number**” **aplikacije** koja je generirala dokument
    - build number = “serijski broj” aplikacije
    - kod prevođenja (kompajliranja) aplikacije, “build number” se povećava
  - Korisno u slučaju kad postoji **sumnja** da je netko **mijenjao vrijeme** nastanka dokumenta

meta.xml – example

```
<?xml version="1.0" encoding="UTF-8"?>
<office:document-meta xmlns:office="...">
        <office:meta>
                <meta:generator>OpenOffice.org/2.4$Unix
OpenOffice.org_project/680m17$Build-9310</meta:generator>
                <dc:title>Create A Template</dc:title>
                <meta:initial-creator>Kristian</meta:initial-creator>
                <meta:creation-date>2008-07-05T22:19:43</meta:creation-date>
                <dc:creator>Kristian</dc:creator>
                <dc:date>2008-08-21T22:19:08</dc:date>
                <dc:language>en-US</dc:language>
                <meta:editing-cycles>3</meta:editing-cycles>
                <meta:editing-duration>PT13M5S</meta:editing-duration>
                ...
                <meta:user-defined meta:name="Info 1"/>
                ...
                <meta:document-statistic meta:table-count="0" meta:image-count="20„
 meta:object-count="0" meta:page-count="6" meta:paragraph-count="77"
meta:word-count="1340" meta:character-count="7790"/>
        </office:meta>
</office:document-meta>
```

PDF - Portable Document Format

- **Najraširenij**i format
  za prenošenje digitalnih dokumenata **nepromjenjivog izgleda**
  - Izgled i raspored sadržaja u dokumentu
    je uvijek isti na svim platformama
- U osnovi sadrži
  - niz PostScript naredbi koje određuju izgled,
  - te dodatne fontove i grafičke elemente
    koje su potrebne za prikaz dokumenta
- Dodatno
  - Omogućuje i interaktivne elemente
    kao checkbox gumbe i polja za unos
  - U novije vrijeme sadrži i
    JavaScript, Flash, poveznice na vanjski sadržaj i drugo...
    - Napomena:
      upravo zbog ovoga se **PDF** dokumenti često koriste
      za **napad na osobna računala**

PDF metadata

- PDF dokumenti sadrže dva tipa metapodataka
  - Document Information Directory
    - Sadrži ***key->value*** parove s
      - » Informacijama o autoru,
      - » naslov dokumenta,
      - » te vremena stvaranja i izmjene dokumenta
  - Extensible Metadata Platform (XMP)
    - Popularna metoda za pohranu metapodataka
    - Često se koristi u formatima za pohranu slika

Example: PDF metadata

```
1 0 obj
<<
/Creator (NameOfOriginalFile.doc - Microsoft Word)
/CreationDate (D:20071001144610Z)
/Title (NameOfOriginalFile.doc)
/Author (UserName)
/Producer (Acrobat PDFWriter 5.0 for Windows NT)
/ModDate (D:20071004091629-05'00')
>>
endobj
```

Zanimljivosti PDF metadata

- ***Creator/Producer*** oznaka je posebno zanimljiva kod analize zloćudnih PDF dokumenata
  - kada se **zloćudni PDF** dokumenti stvaraju putem alata,
  - ti alati često stavljaju **svoj naziv** u ove oznake
- Povijest izmjena
  - Neki PDF alati čuvaju povijest izmjena dokumenta
    - zanimljivo kod dokazivanja autentičnosti
    - ili u slučajevima kada novije verzije dokumenta sadrže manje informacija od starih
- Prilikom pretvaranja datoteka
  - iz originalnog oblika (npr. DOCX) u PDF oblik
  - metapodaci obično ostaju zadržani!
- Svi umetnuti objekti (kao i kod MS Office-a)
    - slike, …
  - također sadrže svoje metapodatke koji su zadržani

Slike

- Jednostavan koncept koji podrazumijeva
  - podatke koje se mogu prikazati (renderirati) kao grafika
- Velik je broj raznih formata
  s različitim svojstvima i primjenama
  - PNG, GIF, JPEG, TIFF …

Metapodaci kod slika

Tri najčešća formata za pohranu metapodataka u slikama:

- EXIF (Exchangable Image File Format)
  - Napravljen s ciljem pohrane metapodataka o uređaju (kameri) koji se koristio za izradu fotografije
  - Sastoji se od niza ***key->value*** parova
- IPTC (Information Interchange Model)
  - Napravljen s ciljem pohrane informacija u slike koje se koriste u novinama i medijskim kućama, te drugim industrijama koje proizvode digitalne slike za medije
- XMP (eXtensible Metadata Platform)
  - Razvila ga organizacija Adobe 2001. godine. Slične je strukture i izgleda kao EXIF.
  - Iako je namijenjen za slike, ponekad se koristi i u drugim formatima (npr. PDF)

EXIF metadata

- oprema kojom je fotografija snimljena
  - aparat (HW i SW)
  - objektiv
  - svojstva senzora
- fotografski parametri
  - žarišna duljina
  - ekspozicija
  - otvor objektiva
  - bljeskalica
  - orijentacija fotoaparata
  - prostor boje
- datum i vrijeme snimanja
- geografske koordinate

Tag	Value
Manufacturer	CASIO
Model	QV-4000
Orientation (rotation)	top - left [8 possible values[21]]
Software	Ver1.01
Date and Time	2003:08:11 16:45:32
YCbCr Positioning	centered
Compression	JPEG compression
x-Resolution	72.00
y-Resolution	72.00
Resolution Unit	Inch
Exposure Time	1/659 sec.
FNumber	f/4.0
ExposureProgram	Normal program
Exif Version	Exif Version 2.1
Date and Time (original)	2003:08:11 16:45:32
Date and Time (digitized)	2003:08:11 16:45:32
ComponentsConfiguration	Y Cb Cr -
Compressed Bits per Pixel	4.01
Exposure Bias	0.0
MaxApertureValue	2.00
Metering Mode	Pattern
Flash	Flash did not fire.
Focal Length	20.1 mm
MakerNote	432 bytes unknown data
FlashPixVersion	FlashPix Version 1.0
Color Space	sRGB
PixelXDimension	2240
PixelYDimension	1680
File Source	DSC
InteroperabilityIndex	R98
InteroperabilityVersion	(null)

Alati

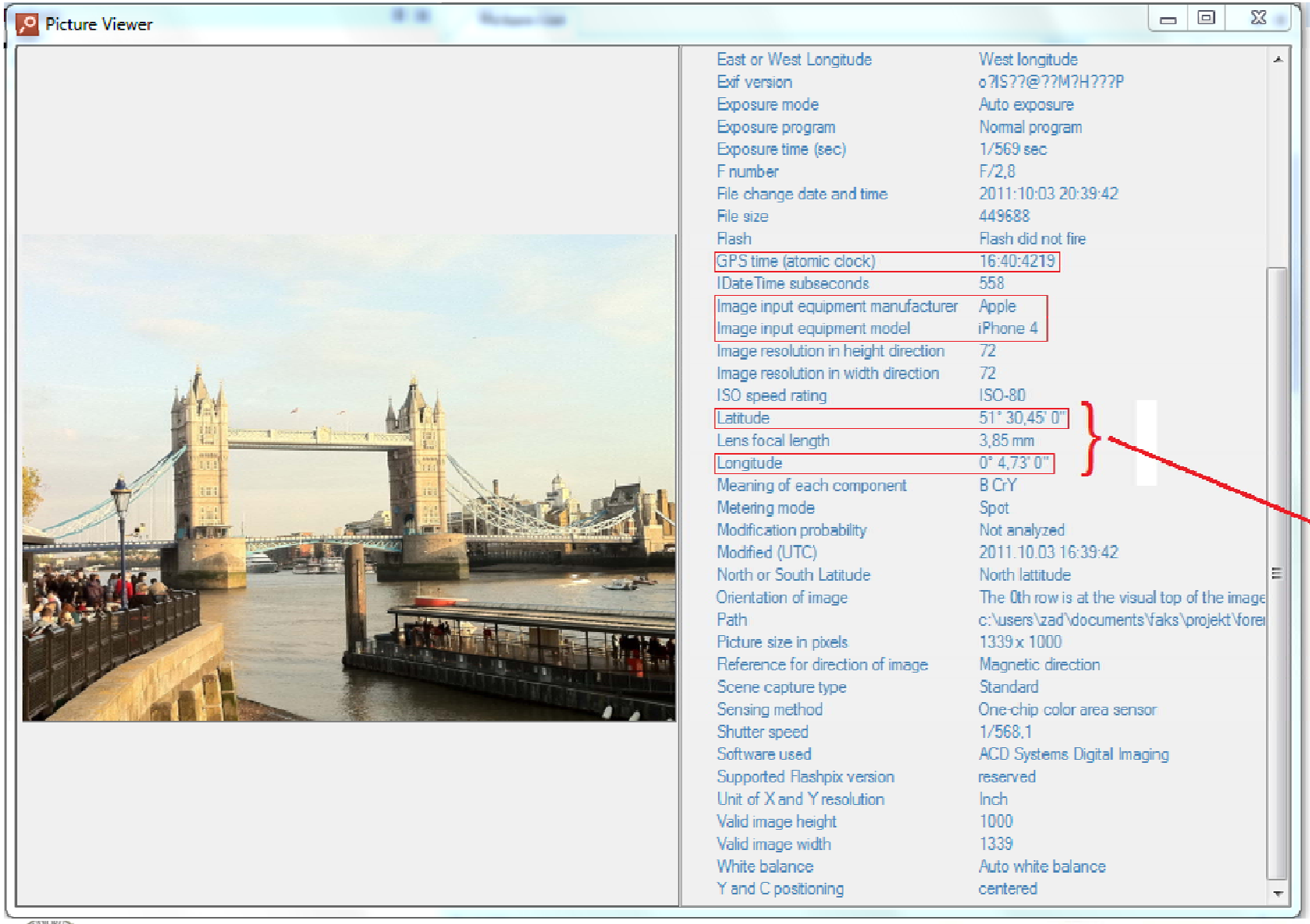
- za ekstrakciju metapodataka kod fotografija:
  - ***Exiftool***
    - općenito *najpopularniji/poznatiji*
  - ***exiv2***
  - ***Hachoir-metadata***

Exiftool – primjer

"exiftool(-k).exe" "C:\Users\kiki\Pictures\2010-06\photo0014(1).jpg

ExifTool Version Number : 10.02
File Name : photo0014(1).jpg
Directory : C:/Users/kiki/Pictures/2010-06
File Size : 507 kB
File Modification Date/Time : 2010:09:17 18:31:46+02:00
File Access Date/Time : 2012:07:16 18:47:20+02:00
File Creation Date/Time : 2010:09:17 18:31:46+02:00
File Permissions : rw-rw-rw-
File Type : JPEG
File Type Extension : jpg
MIME Type : image/jpeg
Exif Byte Order : Little-endian (Intel, II)
Make : Nokia
Camera Model Name : X3-00
X Resolution : 300
Y Resolution : 300
Resolution Unit : inches
Software : V 03.60
Y Cb Cr Positioning : Centered
Exif Version : 0220
Date/Time Original : 2010:06:04 19:20:24
Create Date : 2010:06:04 19:20:24
Components Configuration : Y, Cb, Cr, -
Flashpix Version : 0100
Color Space : sRGB
Exif Image Width : 1536
Exif Image Height : 2048
Custom Rendered : Normal
Exposure Mode : Auto
White Balance : Auto
Digital Zoom Ratio : 1
Scene Capture Type : Standard
Compression : JPEG (old-style)
Thumbnail Offset : 462
Thumbnail Length : 4966
Image Width : 1536
Image Height : 2048
Encoding Process : Baseline DCT, Huffman coding
Bits Per Sample : 8
Color Components : 3
Y Cb Cr Sub Sampling : YCbCr4:4:0 (1 2)
Image Size : 1536x2048
Megapixels : 3.1
Thumbnail Image : (Binary data 4966 bytes, use -b option to extract)

EXIF – Primjer koordinata



Audio formati

- Datoteke koje omogućuju reprodukciju zvuka (ukoliko se ispravno dekodiraju)
  - WAV (Waveform Audio File Format)
    - Razvili su ga IBM i Microsoft za PC, često sarži XMP metapodatke – moguće ih je ekstrahirati pomoću ***hachoir-metadata*** alata
  - MPEG-3/MP3 (Moving Picture Experts Group)
    - Može sadržavati metapodatke u dva formata:
      - ID3v1
        - » Oznake su limitirane na niz od 128 okteta, a dodaju se na kraj MP3 datoteke
      - ID3v2
        - » Oznake su proširene na dodatnih 227 okteta, a dodaju se neposredno prije ID3v1 oznaka
  - Alati ***exiftool*** i ***hachoir-metadata*** mogu ekstrahirati ove metapodatke

EXIF specification for **RIFF** in **WAV**

Tag	Value
Encoding	Microsoft PCM
Num Channels	1
Sample Rate	7872
Avg Bytes Per Sec	7872
Bits Per Sample	8
Date Created	2005:08:08
Exif Version	0220
Related Image File	IMGP1149.JPG
Time Created	16:23:35
Make	PENTAX Corporation
Model	PENTAX Optio WP
MakerNote	(2064 bytes of data)

Video formati

- Poznati formati:
  - MPEG-1 i MPEG-2
    - Ovi formati ne sadrže metapodatke
  - MPEG-4 (MP4)
    - Mogu sadržavati ID3 oznake (poput MP3), te dodatne metapodatke specifične za MP4 datoteke
    - Alat ***AtomicParsley*** se može koristiti za ekstrakciju metapodataka
  - AVI
    - Sadrži metapodatke isto kao WAV format
    - Mogu se ekstrahirati ***hachoir-metadata*** alatom

PRIVREMENE INFORMACIJE

Privremene informacije

- Interaktivni alati za stvaranje dokumenata
  - poput **office alata**
- mogu **pohraniti informacije** u dokumentu koje su
  - **stare**
  - prolazni / **privremeni**
  - **ostavljeni** zbog programskih pogrešaka
- koje su "**nevidljive**" običnim korisnicima
- ali koje **mogu biti otkrivene** posebnim alatima
- a otkrivaju potencijalno **osjetljive informacije**
  - Starije verzija teksta
  - Druge autore
  - Lozinke
  - …

GARBAGE PROSTOR

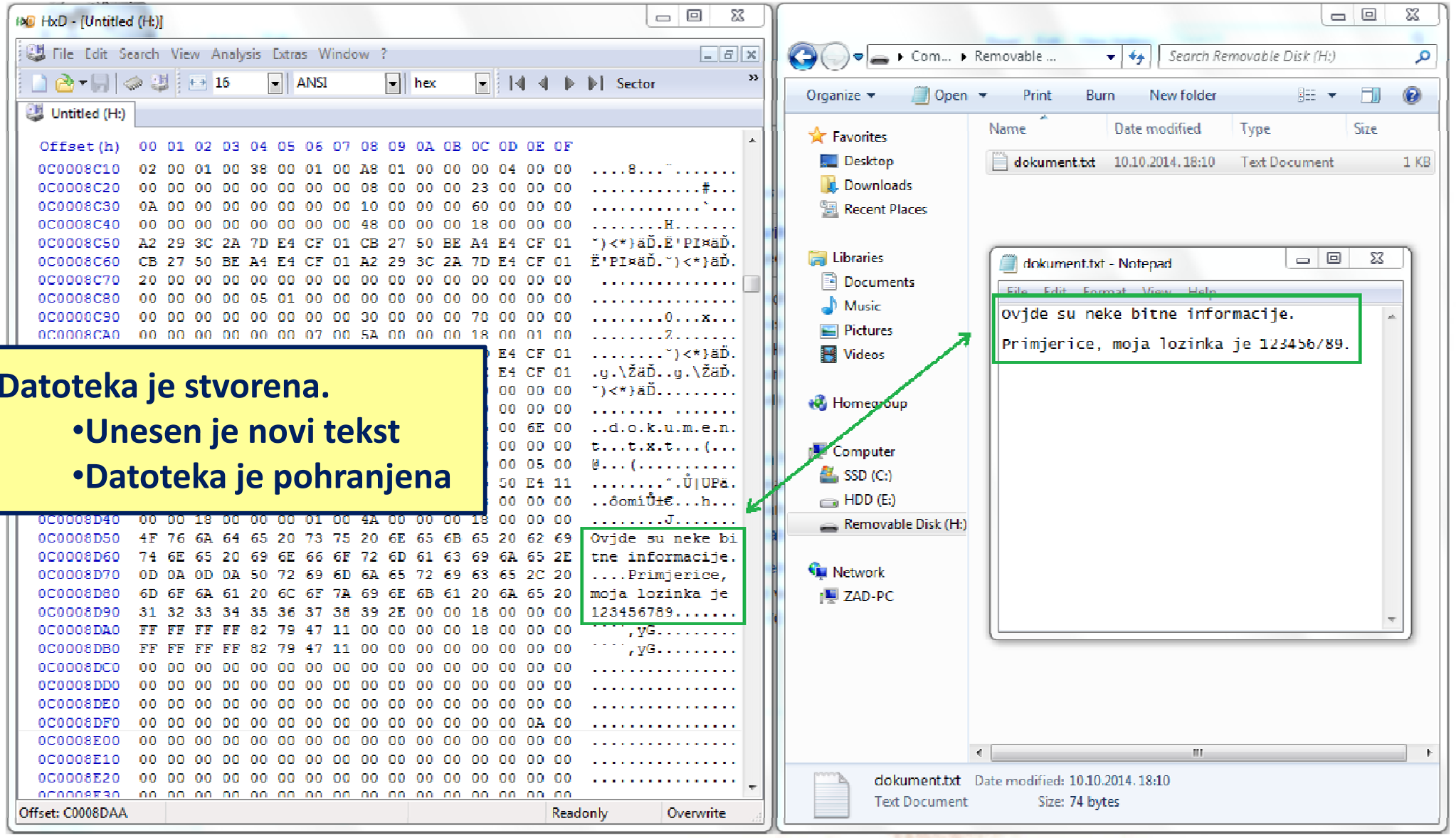
Garbage prostor

- Prostor na diskovima je podijeljen u blokove
  - Blokovi su veličine 512, 1024, 2048... bytea
  - Datoteke zauzimaju **diskretnu** količinu blokova na disku
    - Znači da **ne mogu** zauzeti, na primjer, **pola bloka**
- Na primjer:
  - Veličina bloka: 4096 bytea
  - Datoteka ***document.txt*** ima **samo 100** bytea informacija
  - Ali – na disku **zauzima svih 4096** bytea (1 cijeli blok)
  - PITANJE: Što se nalazi u preostalih 3996 byteova?
    - "Prazan" prostor – koji može sadržavati bitne informacije!
  - U tom prostoru nisu podaci koji pripadaju datoteci *document.txt*
  - već podaci **preostali** od datoteke koja **je obrisana**,
    a prije je **koristila isti blok** na disku

Garbage prostor – primjer

- Koristeći **program** koji **čita**
  "čiste" podatke (byteove) **s diska**
  možemo vidjeti što se skriva u tom praznom prostoru
- Kreirat ćemo tekstualnu datoteku te ćemo ju:
  - **Izmijeniti**
  - **Izbrisati**
  - **Izbrisati** i **stvoriti** novu datoteku
- **Nakon svakog koraka** ćemo **provjeriti**
  što se nalazi u tom "**praznom**" prostoru

Garbage prostor – demonstracija 1



Datoteka je stvorena.

- **Unesen je novi tekst**
- **Datoteka je pohranjena**

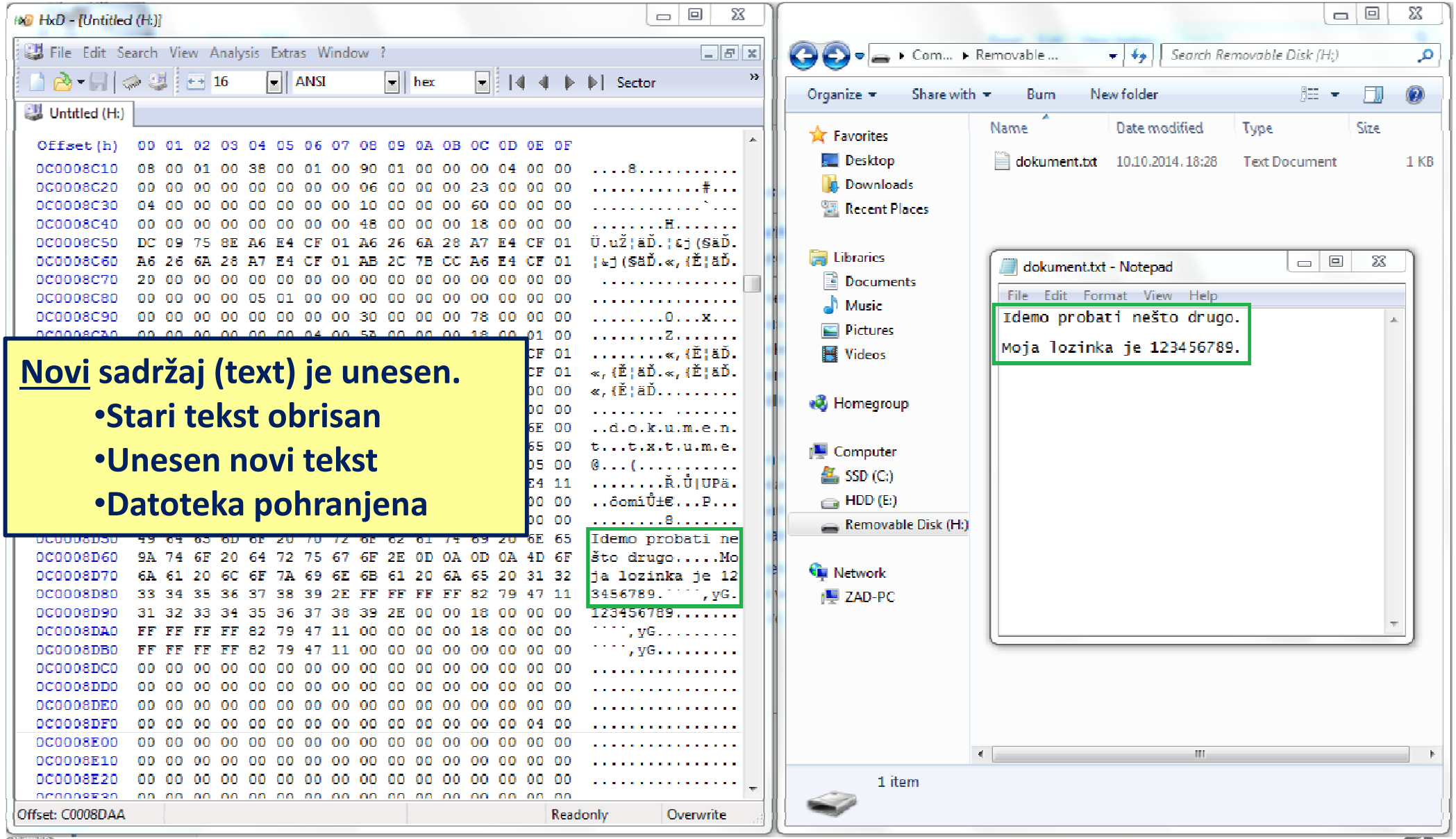
Garbage prostor – demonstracija 2

Sadržaj datoteke je izmijenjen.

- **Stari tekst obrisan**
- **Unesen novi tekst**
- **Datoteka pohranjena**

Sadržaj ostao od prethodnog uređivanja!

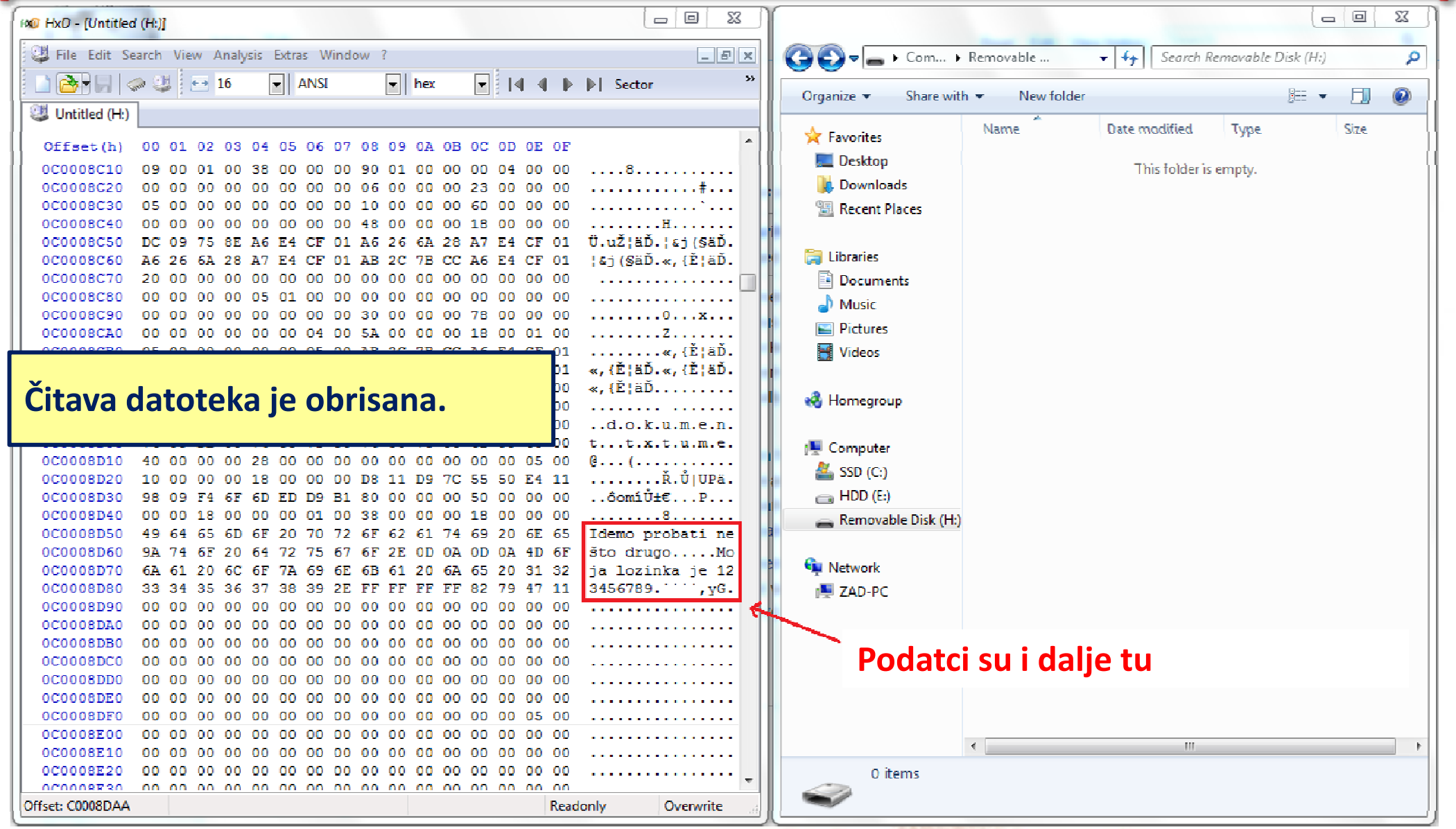
Garbage prostor – demonstracija 3



Novi sadržaj (text) je unesen.

- Stari tekst obrisan
- Unesen novi tekst
- Datoteka pohranjena

Garbage prostor – demonstracija 4



Čitava datoteka je obrisana.

Podatci su i dalje tu

Garbage prostor – demonstracija 5

Nova datoteka

- Je stvorena
- Unesen novi tekst
- Datoteka pohranjena

Garbage prostor – demonstracija 6

Nova datoteka otvorena u Hex alatu

Stari podatci i dalje tu!

Za analizu dokumenata koristimo:

- Ekstenzije datoteka
  - Brza identifikacija vrste sadržaja
  - Nepouzdana jer se lako manipulira
- “Magic numbers”
  - pouzdanija
- Metapodatke
  - Dodatne informacije o sadržaju
- “Garbage” prostor
  - Informacija o drugim datotekama


RacFor.zesoi.fer.hr
RacFor@zesoi.fer.hr