Searching for relationships between forensic artifacts and their attributes

Women in Forensic Computing Workshop – March 28, 2022

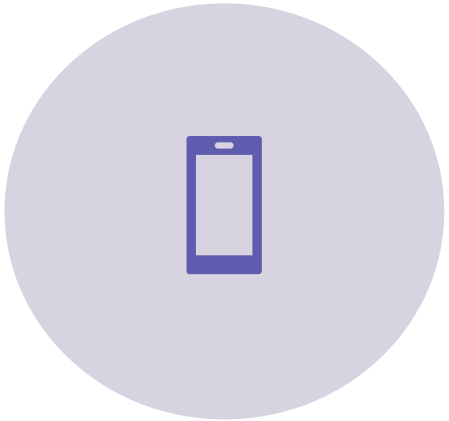
Eva Marková
Pavol Jozef Šafárik Univesrity in Košice, Faculty of Science
E-mail: eva.markova@upjs.sk

- PhD. Student at Pavol Jozef Šafárik University, Košice, Slovakia
- Member of CSIRT-UPJS team – first accredited academic CSIRT team in Slovakia
  - Solving security incidents
  - Education
  - Research
  - Spreading security awareness

https://csirt.upjs.sk/#/en/

Motivation

An increasing number of cyberattacks



A lot of alerts from monitoring devices

Better orientation in the amount of data that digital forensics deals with

Finding interesting records in the data

Description of model case

- The Stolen Szechuan Sauce
- Used to teach digital forensics, incident response and threat hunting
- Analysis of unauthorized intrusion into the network of company CITADEL

- Task - identify whether
  - malicious applications have been installed on the system
  - any information has been created, modified, or deleted in the system
  - data has been leaked

Preprocessing of data

- Use of the disk image of the subject server (DC01-E01) as input

1. Creation of a timeline of the image using Plaso tool
2. Conversion of the plaso file to the l2tcsv format – 17 default fields
3. Division of records according to values in the field source – 11 dataframes (the most important – file, evt, reg)
4. Extraction of additional attributes
5. Aggregation

Filename

```
7603                                                                                                NTFS:\Windows\SysWOW64\ctl3d32.dll
7604                                    NTFS:\Windows\WinSxS\x86_microsoft-windows-ctl3d32_31bf3856ad364e35_6.3.9600.16384_none_5ec8aaef6fba350f\ctl3d32.dll
7605                                                                                                NTFS:\Windows\SysWOW64\stdole32.tlb
7606                                                                                                NTFS:\Windows\System32\stdole32.tlb
7607                      NTFS:\Windows\WinSxS\amd64_microsoft-windows-o..ion-legacy-stdole32_31bf3856ad364e35_6.3.9600.16384_none_d28f0f52ebefe7cd\stdole32.tlb
                                                                                                    ...
1263782    NTFS:\Windows\WinSxS\wow64_microsoft-windows-p..man-pluginworker-v2_31bf3856ad364e35_6.3.9600.16384_none_9289c385fafccf5c\pspluginwkr.dll
1263783                                                                                             NTFS:\Windows\System32\WindowsPowerShell\v1.0\pspluginwkr.dll
1263784    NTFS:\Windows\WinSxS\amd64_microsoft-windows-p..man-pluginworker-v2_31bf3856ad364e35_6.3.9600.16384_none_88351933c69c0d61\pspluginwkr.dll
1263785                                                                                             NTFS:\Windows\SysWOW64\WindowsPowerShell\v1.0\pspluginwkr.dll
1263786    NTFS:\Windows\WinSxS\wow64_microsoft-windows-p..man-pluginworker-v2_31bf3856ad364e35_6.3.9600.16384_none_9289c385fafccf5c\pspluginwkr.dll
Name: filename, Length: 1256180, dtype: object
```

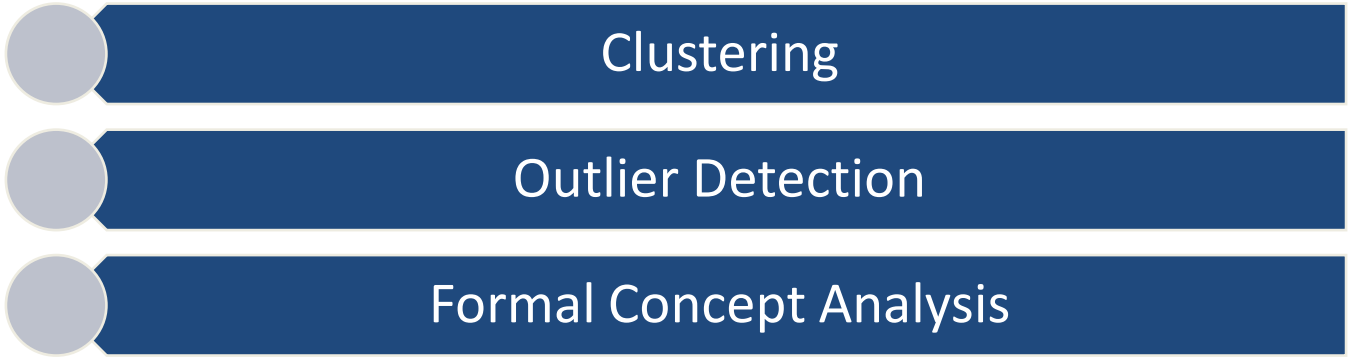
	date	time	timezone	MACB	source	sourcetype	type	user	host	short	desc	version	filename	inode	notes	format	extra
228373	08/22/2013	15:26:10	UTC	...B	FILE	NTFS file stat	Creation Time	-	-	\$MFT 71146-1 $STANDARD_INFORMATION	NTFS:\$MFT File reference: 71146-1 Attribute n...	2	NTFS:\$MFT	0	-	mft	attribute_type: 16; file_system_type: NTFS; is...
866929	09/17/2020	16:49:27	UTC	MACB	FILE	NTFS file stat	Content Modification Time; Creation Time; Last...	-	-	\$MFT 56297-1 $FILE_NAME	NTFS:\$MFT File reference: 56297-1 Attribute n...	2	NTFS:\$MFT	0	-	mft	attribute_type: 48; file_attribute_flags: 32; ...
662807	09/17/2020	16:47:31	UTC	MACB	FILE	NTFS file stat	Content Modification Time; Creation Time; Last...	-	-	\$MFT 85590-2 $FILE_NAME	NTFS:\$MFT File reference: 85590-2 Attribute n...	2	NTFS:\$MFT	0	-	mft	attribute_type: 48; file_attribute_flags: 32; ...
1128467	09/17/2020	17:56:11	UTC	M...	REG	Registry Key	Content Modification Time	-	-	[HKEY_LOCAL_MACHINE\Software\Microsoft\Windows...	[HKEY_LOCAL_MACHINE\Software\Microsoft\Windows...	2	NTFS:\Windows\System32\config\SOFTWARE	78036	-	winreg/winreg_default	sha256_hash: 9ea369b327a9241abd6ed3f2801218cf3...
867216	09/17/2020	16:49:27	UTC	..C.	FILE	NTFS file stat	Metadata Modification Time	-	-	\$MFT 56355-1 $STANDARD_INFORMATION	NTFS:\$MFT File reference: 56355-1 Attribute n...	2	NTFS:\$MFT	0	-	mft	attribute_type: 16; file_system_type: NTFS; is...

Extra

```
539634       file_size: 758; file_system_type: NTFS; is_allocated: True; sha256_hash: 5c2166cf4bb2fb5aee51b6bbcc1bb5b5a2c8bc193c862fc732195e4fbd825c14
987903     file_size: 29397; file_system_type: NTFS; is_allocated: True; sha256_hash: 0627c02f984d62df6597c315b58971ac00ff24babc10511b0e85a587e91fe350
1009580                                                                  attribute_type: 16; file_system_type: NTFS; is_allocated: True
511548                                                                   attribute_type: 16; file_system_type: NTFS; is_allocated: True
681998                                        attribute_type: 48; file_attribute_flags: 268435456; file_system_type: NTFS; is_allocated: True
Name: extra, dtype: object
```

Finding relationship between attributes

- Used records related to filesystem
- 843863 records
- For the purpose of analysis we have used about a quarter of records
- Methods:



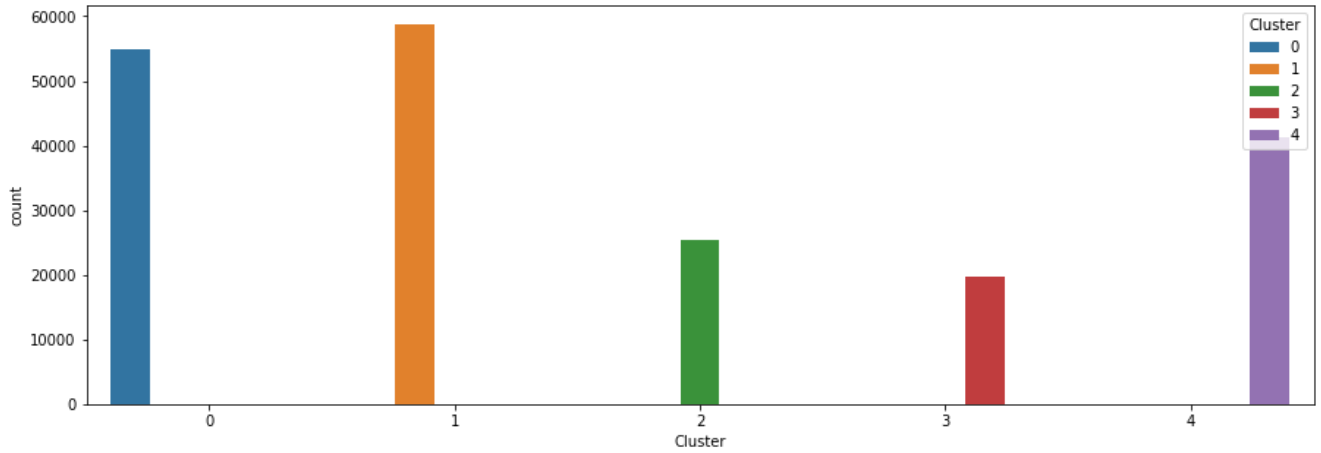
Clustering methods

K-Means – numeric variables

K-Modes – categorical variables

K-Prototypes – both

K-Modes



M	A	C	B	source_file	file_stat	NTFS_file_stat	file_entry_shell...	NTFS_USN_change	name	filef	directory	link	dir_appdata	dir_win	dir_user	dir_other
1	0	0	0	1	1	0	0	0	None	1	0	0	0	1	0	0
0	1	0	1	1	1	0	0	0	None	1	0	0	0	1	0	0
0	1	0	1	1	0	1	0	0	None	0	0	0	0	0	0	1
0	0	0	1	1	0	1	0	0	None	0	0	0	0	0	0	1
1	0	0	0	1	0	1	0	0	None	0	0	0	0	0	0	1

file_exe...	file_gra...	file_doc...	file_ps	file_other	mft	lnk_shell_items	olecf_olecf_autom...	winreg_bagmru...	usnjrnl	is_allocated1	is_allocated0	size_none	size_Q1	size_Q2	size_Q3	size_Q4
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0

Outlier detection

- **Local Outlier Factor**
  - Distance between two boolean vectors: jaccard, kulsinski,...
  - Distance between two numeric vectors: euclidean, minkowski,...
  - Doesn’t work with categorical datasets

- **Oneclass SVM**
  - Kernel - RBF, linear, poly, sigmoid

Example of outliers (29)

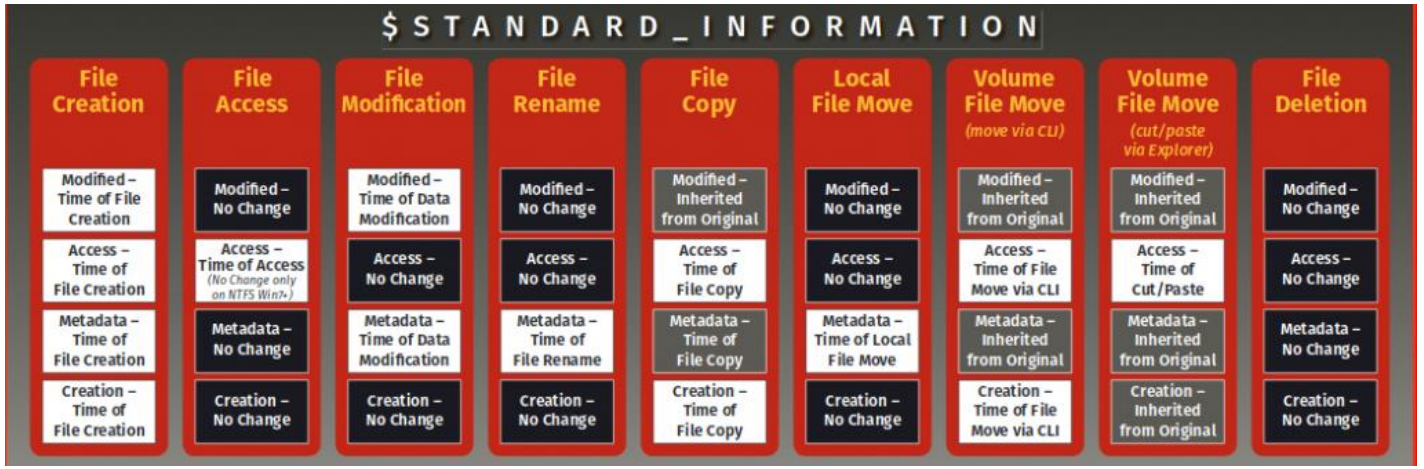
M	A	C	B	...	is_allocated	is_allocated1	is_allocated0	size_none	size_Q1	size_Q2	size_Q3	size_Q4	sha_256
0	1	0	1	...	None	0	0	1	0	0	0	0	653bf59285bb4f514c0b95d3151c529d5678636081778...
1	0	0	0	...	None	0	0	1	0	0	0	0	653bf59285bb4f514c0b95d3151c529d5678636081778...
1	1	0	1	...	None	0	0	1	0	0	0	0	653bf59285bb4f514c0b95d3151c529d5678636081778...
0	0	0	1	...	True	0	0	0	0	1	0	0	691c0a552579e2a312c76ef1ff03029fa5ca4117174df...
0	0	0	1	...	True	0	0	0	0	1	0	0	1b8b4a266b84aebd6f18bb87d679a4b08dac25217b7ed...
0	1	0	0	...	True	0	0	0	0	1	0	0	1b8b4a266b84aebd6f18bb87d679a4b08dac25217b7ed...
1	0	0	0	...	True	0	0	0	0	1	0	0	1b8b4a266b84aebd6f18bb87d679a4b08dac25217b7ed...
0	1	0	0	...	True	0	0	0	0	1	0	0	691c0a552579e2a312c76ef1ff03029fa5ca4117174df...
1	0	0	0	...	True	0	0	0	0	1	0	0	691c0a552579e2a312c76ef1ff03029fa5ca4117174df...

Formal concept analysis

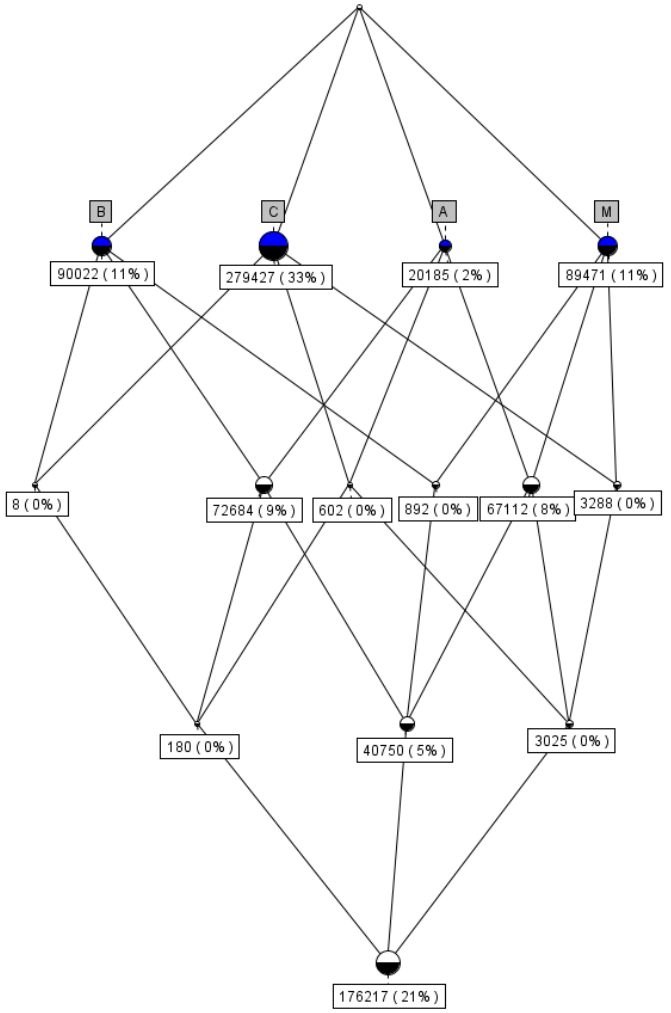
- is a method mainly used for the analysis of data, i.e. for deriving implicit relationships between objects described through a set of attributes on the one hand and these attributes on the other

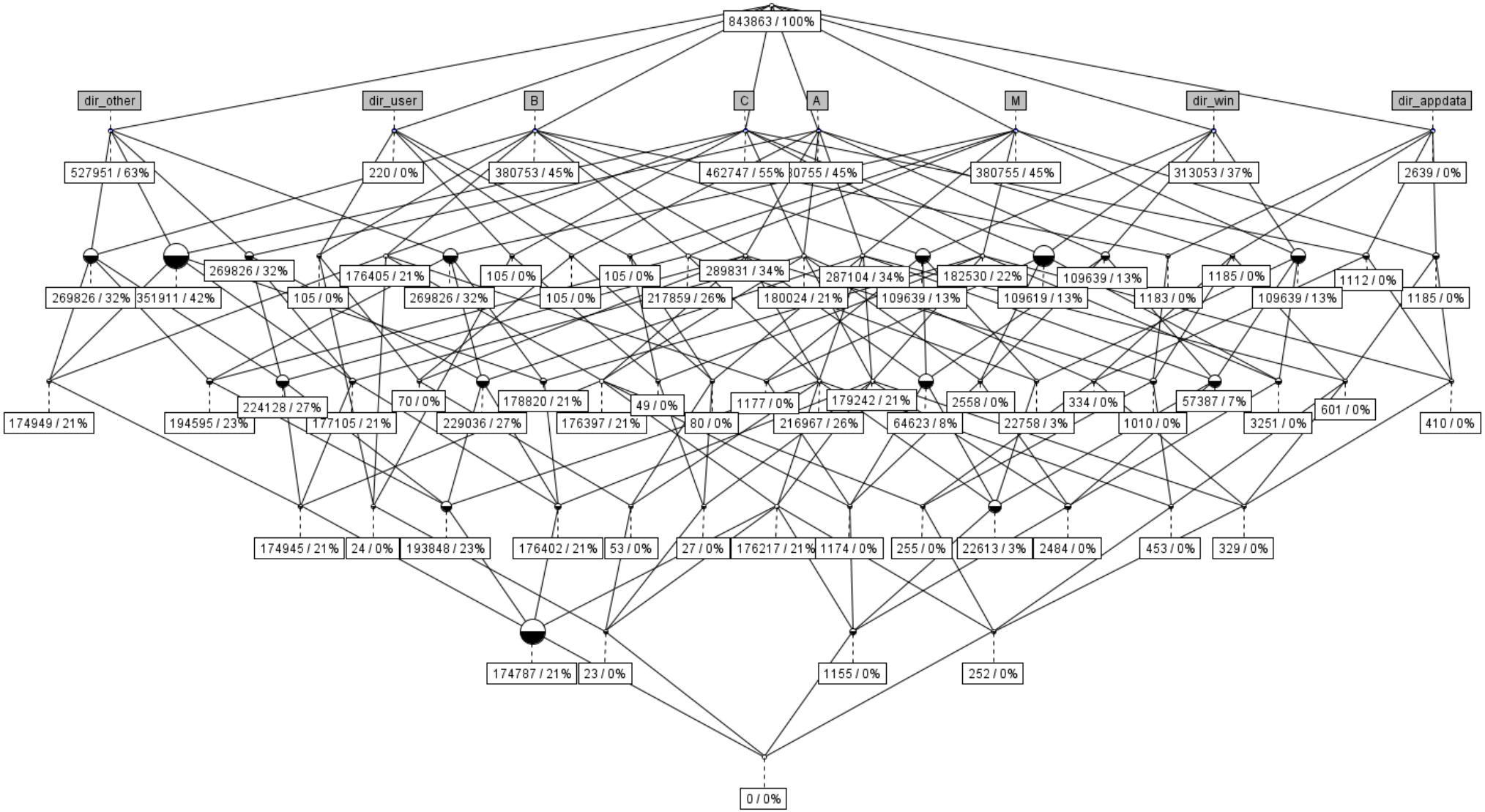
Concept lattice of digital evidence

- Building of a concept lattice relates to a notion of Galois connection
- We can build formal concepts, which are the pairs of extents (subsets of objects) and intents (subsets of attributes)



Source: http://www.jaringankita.com/blog/wp-content/uploads/2020/05/SANS_macb-1024x730.png




843863 / 100%
dir_other
dir_user
B
C
A
M
dir_win
dir_appdata
527951 / 63%
220 / 0%
380753 / 45%
462747 / 55%
380755 / 45%
380755 / 45%
313053 / 37%
2639 / 0%
269826 / 32%
176405 / 21%
105 / 0%
105 / 0%
289831 / 34%
287104 / 34%
182530 / 22%
109639 / 13%
1185 / 0%
1112 / 0%
269826 / 32%
351911 / 42%
105 / 0%
269826 / 32%
105 / 0%
217859 / 26%
180024 / 21%
109639 / 13%
109619 / 13%
1183 / 0%
109639 / 13%
1185 / 0%
174949 / 21%
194595 / 23%
224128 / 27%
177105 / 21%
70 / 0%
229036 / 27%
178820 / 21%
176397 / 21%
49 / 0%
80 / 0%
1177 / 0%
216967 / 26%
179242 / 21%
64623 / 8%
2558 / 0%
22758 / 3%
334 / 0%
1010 / 0%
57387 / 7%
3251 / 0%
601 / 0%
410 / 0%
174945 / 21%
24 / 0%
193848 / 23%
176402 / 21%
53 / 0%
27 / 0%
176217 / 21%
1174 / 0%
255 / 0%
22613 / 3%
2484 / 0%
453 / 0%
329 / 0%
174787 / 21%
23 / 0%
1155 / 0%
252 / 0%
0 / 0%

Association rules of digital evidence

No analysis required

1 < 176217 > M C B =[100%]=> < 176217 > A;

2 < 453 > M B dir_appdata =[100%]=> < 453 > A;

...

But what about „outliers“?

16 < 176217 > M A C B =[99%]=> < 174787 > dir_other;

25 < 24 > C B dir_user =[96%]=> < 23 > M A; (NTUSER.dat)

31 < 27 > M A C dir_user =[85%]=> < 23 > B;

Summary

- **Clustering** - grouping similar records into clusters
  - Consideration: How does it look in different types of incidents?
- **Outliers** - identification of interesting records
  - Problem: We can't use different types of distance
- **Concept lattices** - identification of relevant attributes (set of attributes)
  - Consideration: What if it is done through more cases, are there similarities in relationships?

Conclusion

- Clustering methods, outlier detection and formal concept analysis gave us interesting intermediate data
- Main goals:
  - Finding a suitable set of attributes
  - Finding a suitable method for selecting interesting records
  - Finding or creation of a suitable dataset

- Any insights and thoughts how to continue in our research are **welcome** ☺

Thank you for your attention

E-mail: eva.markova@upjs.sk