

There are **new waves of threats** coming out against your organization. Partner with the **only local source** to address all these threats through a true **security lifecycle**.



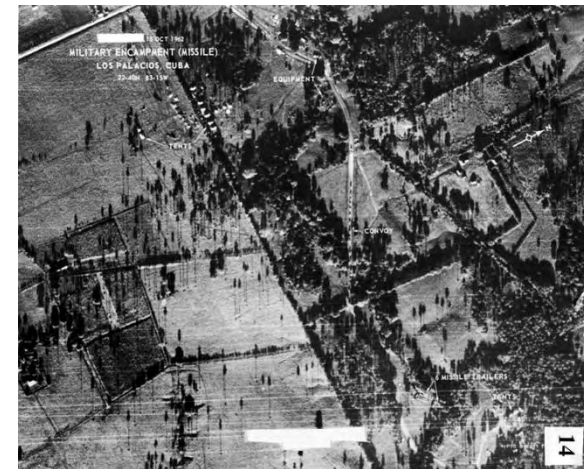
Dangerous Minds: The Art of Guerrilla Data Mining

Mark Ryan del Moral Talabis

- **“Security Analytics”:**
 - **Concept of using data mining and AI in security**
 - **Presented techniques and theories that we could use**
- **This Talk:**
 - **Move from theory to practical applications**
 - **Provide scenarios and examples to leverage these techniques for your research**



- **Traditional warfare vs. Information Security**
 - **Very similar**
 - **Reconnaissance, information gathering, and espionage play an important part in battle tactics**



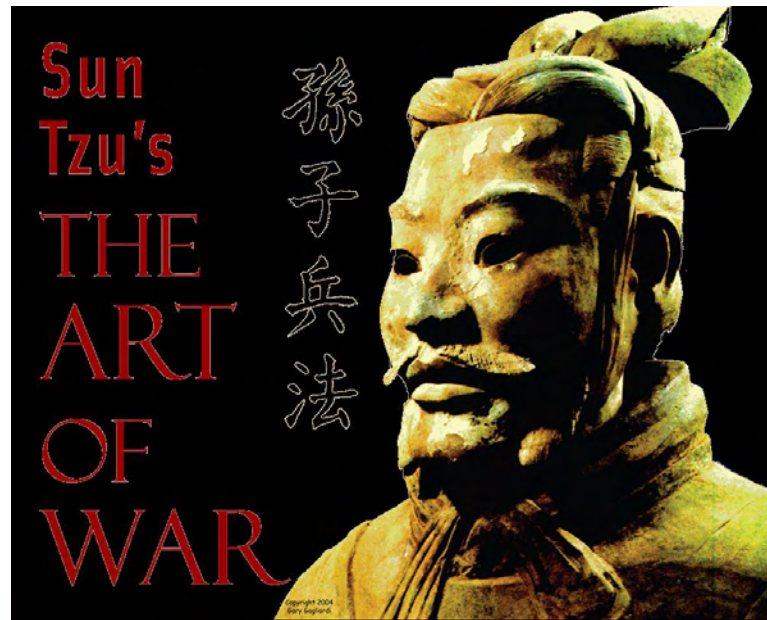
- **Anyone watched 300?**
 - **Spartans:** they knew and understood the terrain
 - **Persians:** They did not win because of overwhelming numbers, they actually won because someone told them about a hidden pass.



- **In information security:**
 - Not only in “hacking” systems
 - The more information you have, you’ll have a better chance to protect you organization
 - Drafting good policies and procedures as well as picking the correct tools and techniques based on the information that you have.



- **It is said that if you know your enemies and know yourself, you will not be imperiled in a hundred battles - *Sun Tzu***



- **Information Warfare**
 - **The use and management of information in pursuit of a competitive advantage over an opponent**
 - **Information are just 1's and 0's if not used properly**
 - **Analysis makes information meaningful - INTELLIGENCE**



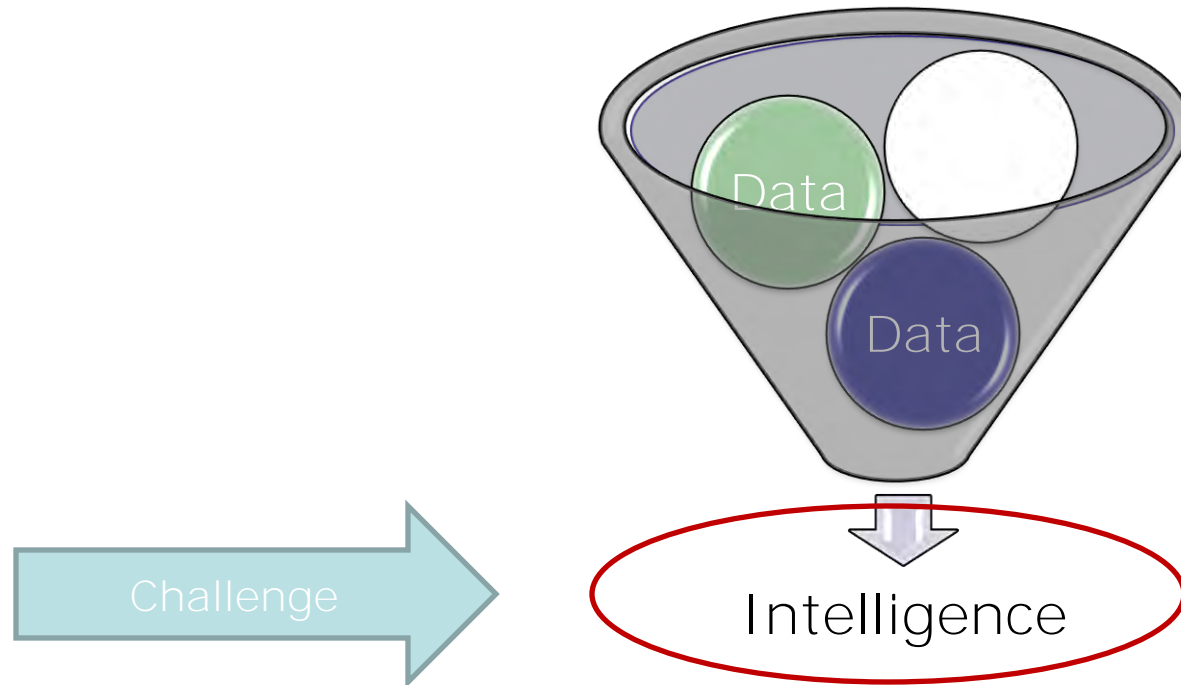
- **People who are into the Information Warfare Business:**
 - CIA
 - FBI
 - NSA
 - Information Awareness Office
 - Foreign Governments



- **Government Projects:**
 - ECHELON
 - TALON
 - ADVISE
 - MATRIX
 - Able Danger
- **Large endeavors!**



- **Amount of data: there's just too much**
- **Resources: way too little**

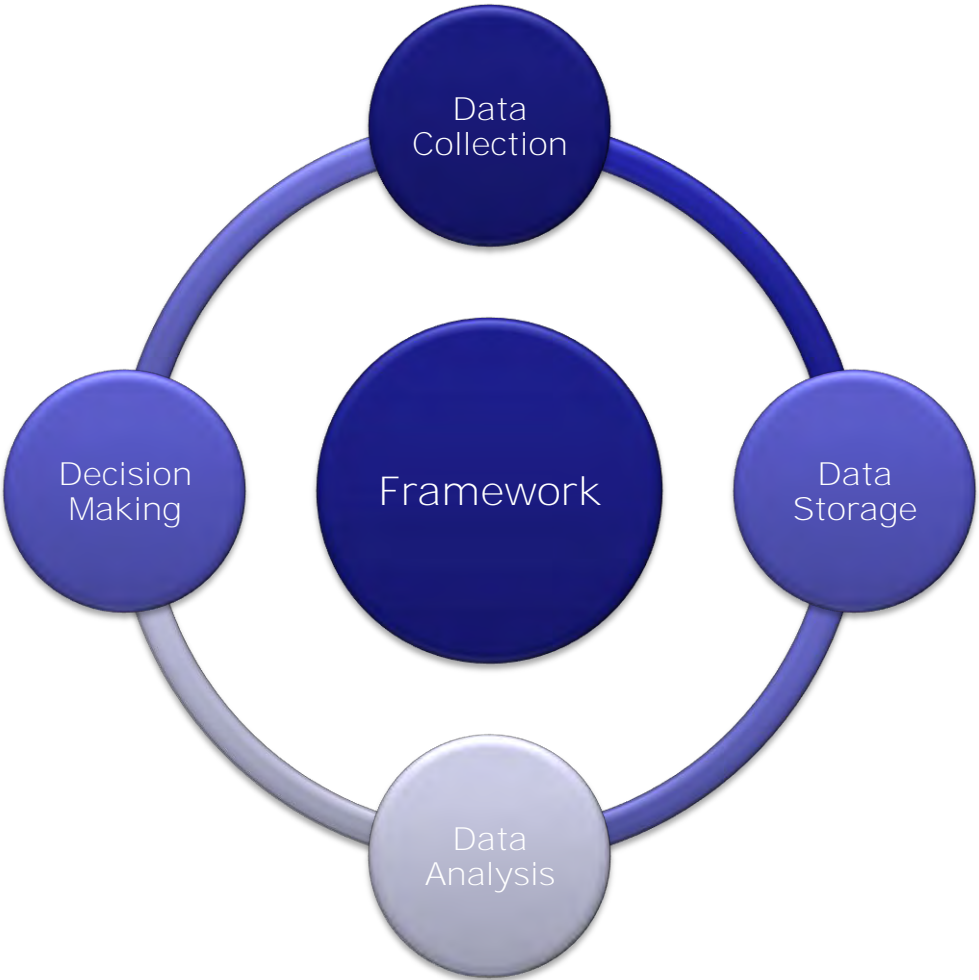




- **Veritas is latin for “Truth”**
- **The Veritas Project**
 - **Modeled in the same general threat intelligence premise**
 - **Primarily based on community sharing approach and using tools, technologies, and techniques that are freely available.**
 - **Hawaii Honeynet Project and Secure-DNA**







- **Sources of Data**

- **Depends on what you want to research**

- **Forums**
 - **Bulletins**
 - **Chat logs**
 - **News**
 - **Articles**
 - **Blogs**
 - **Word documents**

- **The more you can gather, the better results**

- **It's not as easy, unless you're Google**



- **Information can be stored in:**
 - **Relational databases**
 - **Flat files**
- **Possibly the easiest part of all this**

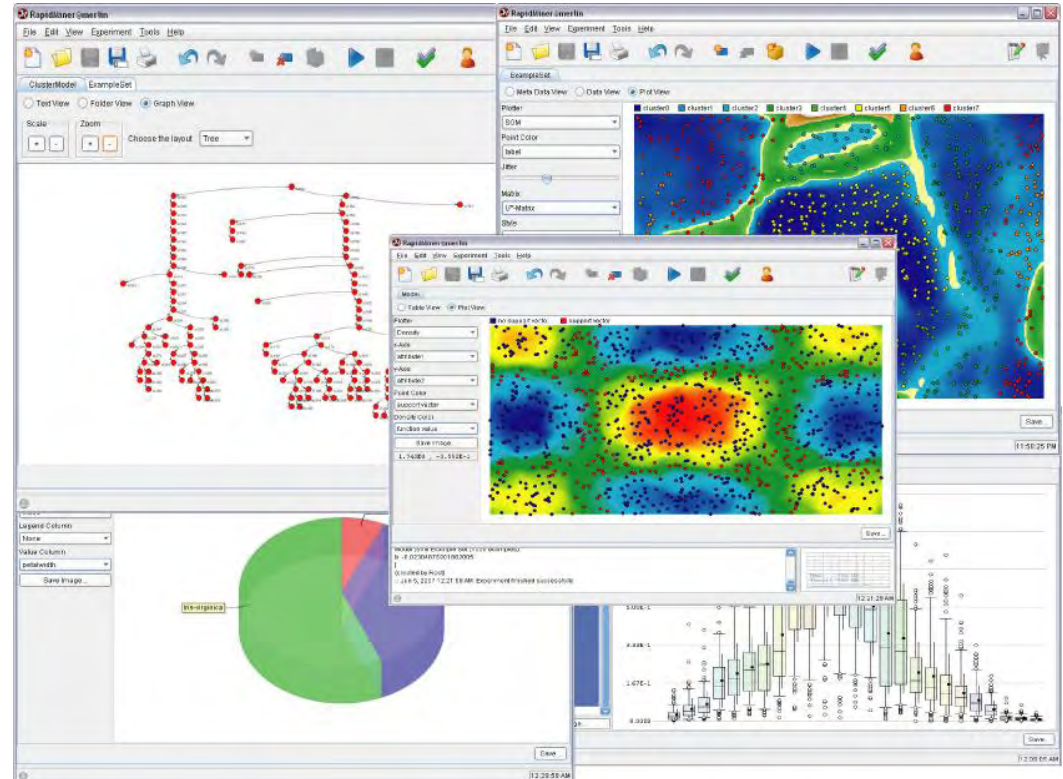


- **Possible the most important aspect of the framework**
 - **Crunching large amount of data.**
 - **Making data and information meaningful**
- **Some Data Mining and Artificial Intelligence Concepts**
 - **K-Means**
 - **Neural Networks**
 - **SVM**
 - **A lot more**
- **Not too easy but there are a lot of tools out there**



▪ Some very useful tools that are free

- Text Garden
- Ontogen
- Weka
- Rapid Miner
- Tanagra
- Orange
- MEAD



- **Why do we need humans?**
- **Interpretation of Results and Analysis = Intelligence**





- **Let's look at the scenarios that you can use as templates for your own research**

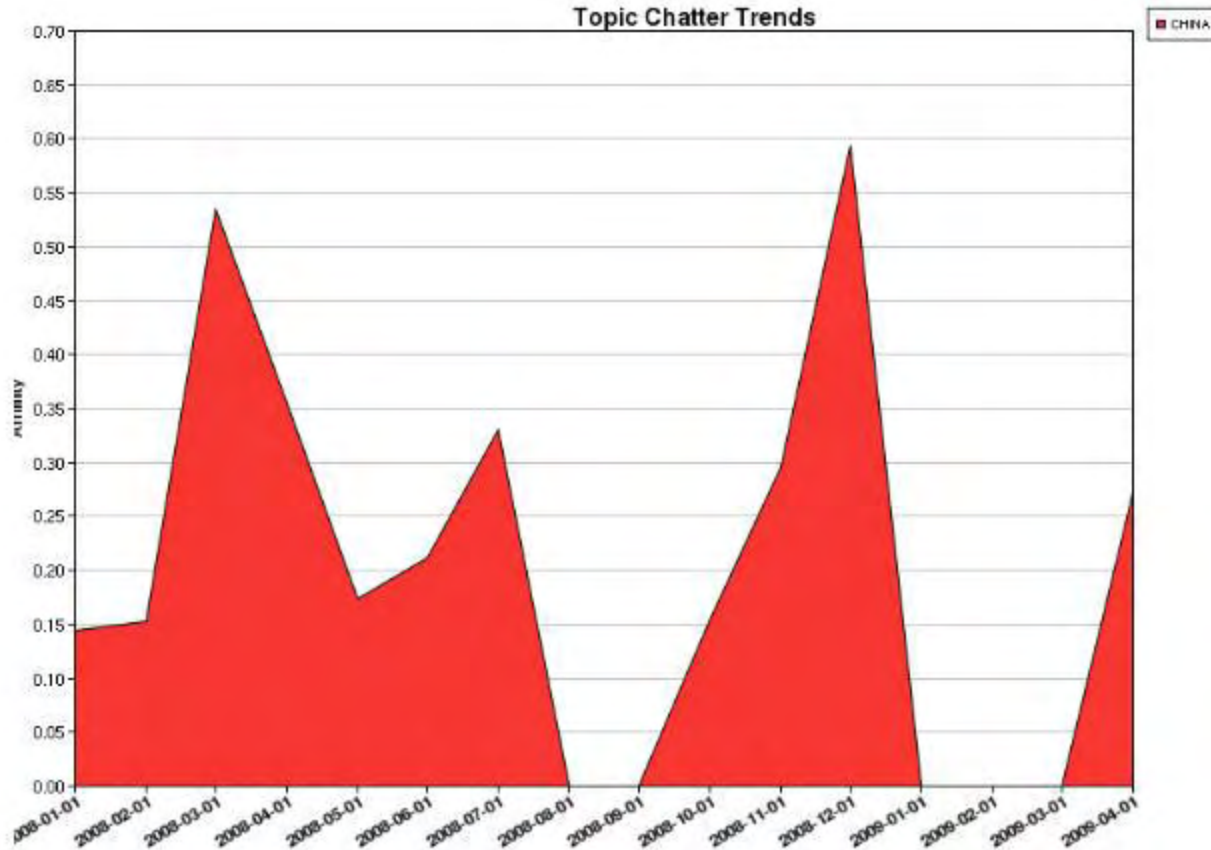


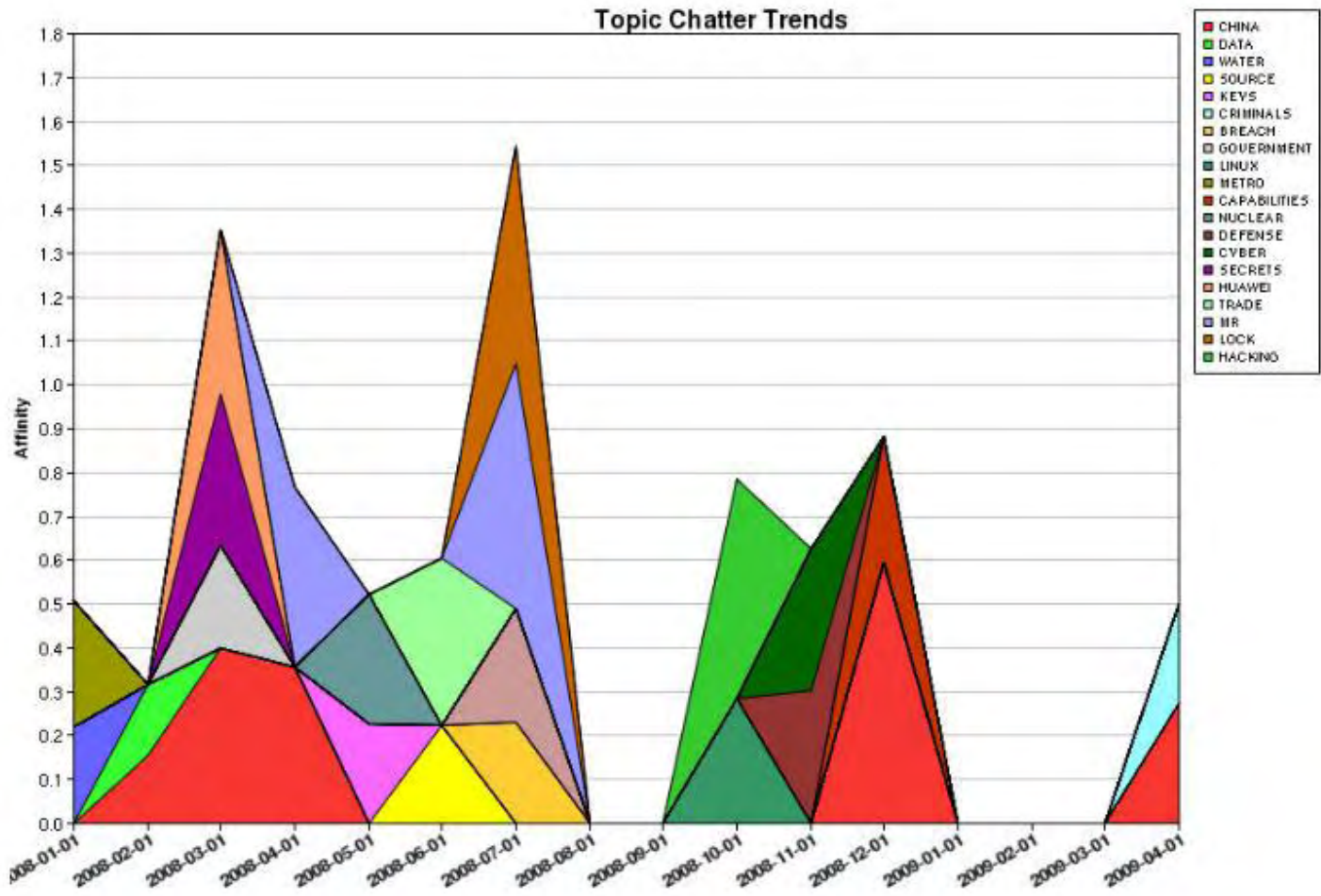
- **Examples**
 - Trends Research
 - Malware Taxonomy
 - Monitoring – Persons of Interest
 - Corporate Intelligence - Strategy
 - Opinion Polls – What people are thinking about

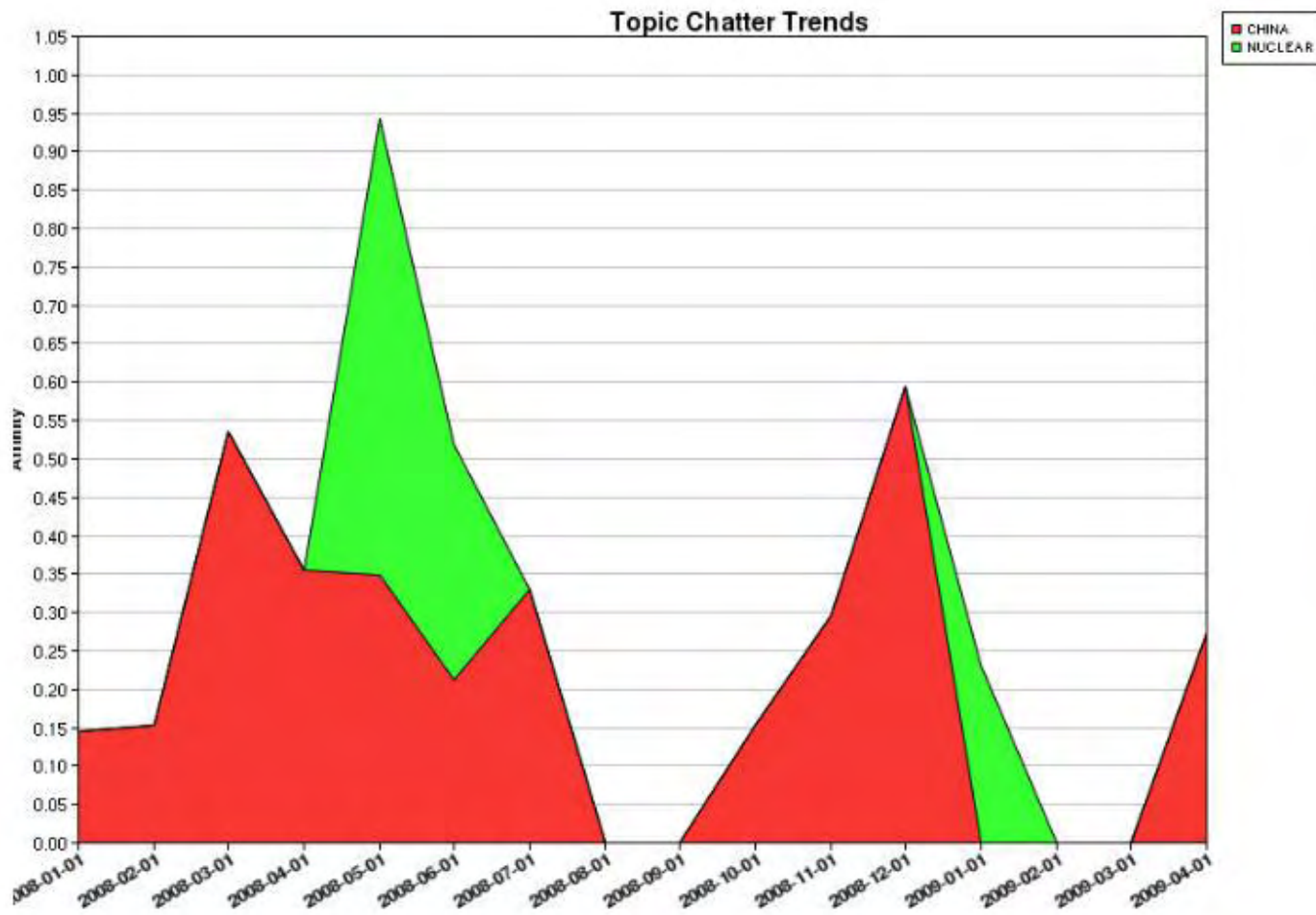


- **Track increases in chatter across time**
 - Gives researchers focus
 - Find relationships between topics
- **Framework**
 - **Data Collection**
 - Crawlers (News articles, Forums)
 - **Data Storage**
 - MySQL
 - **Analysis**
 - Text Garden (html2txt, txt2bow, bowkmeans)
 - **Decision Making**
 - Me!







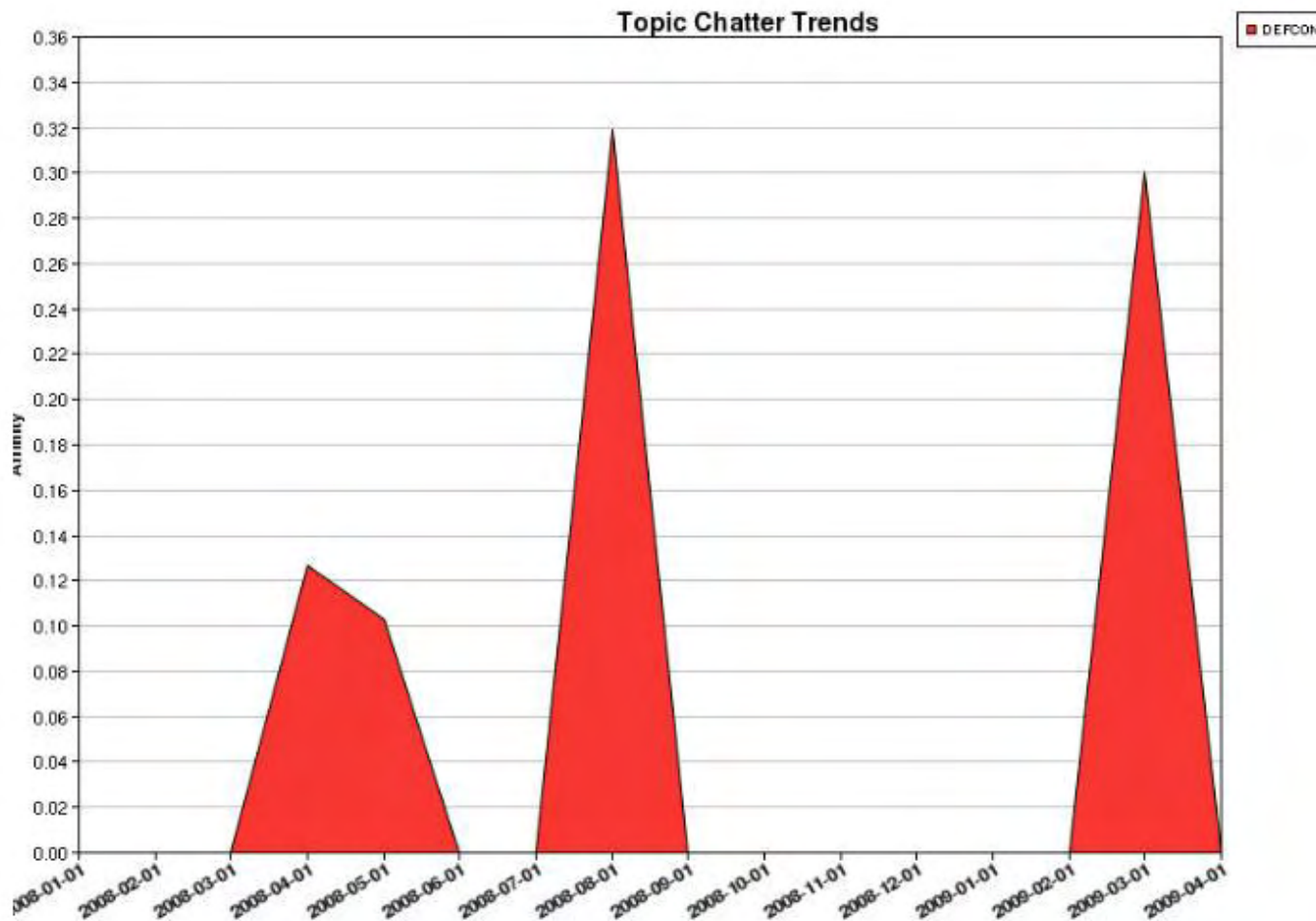


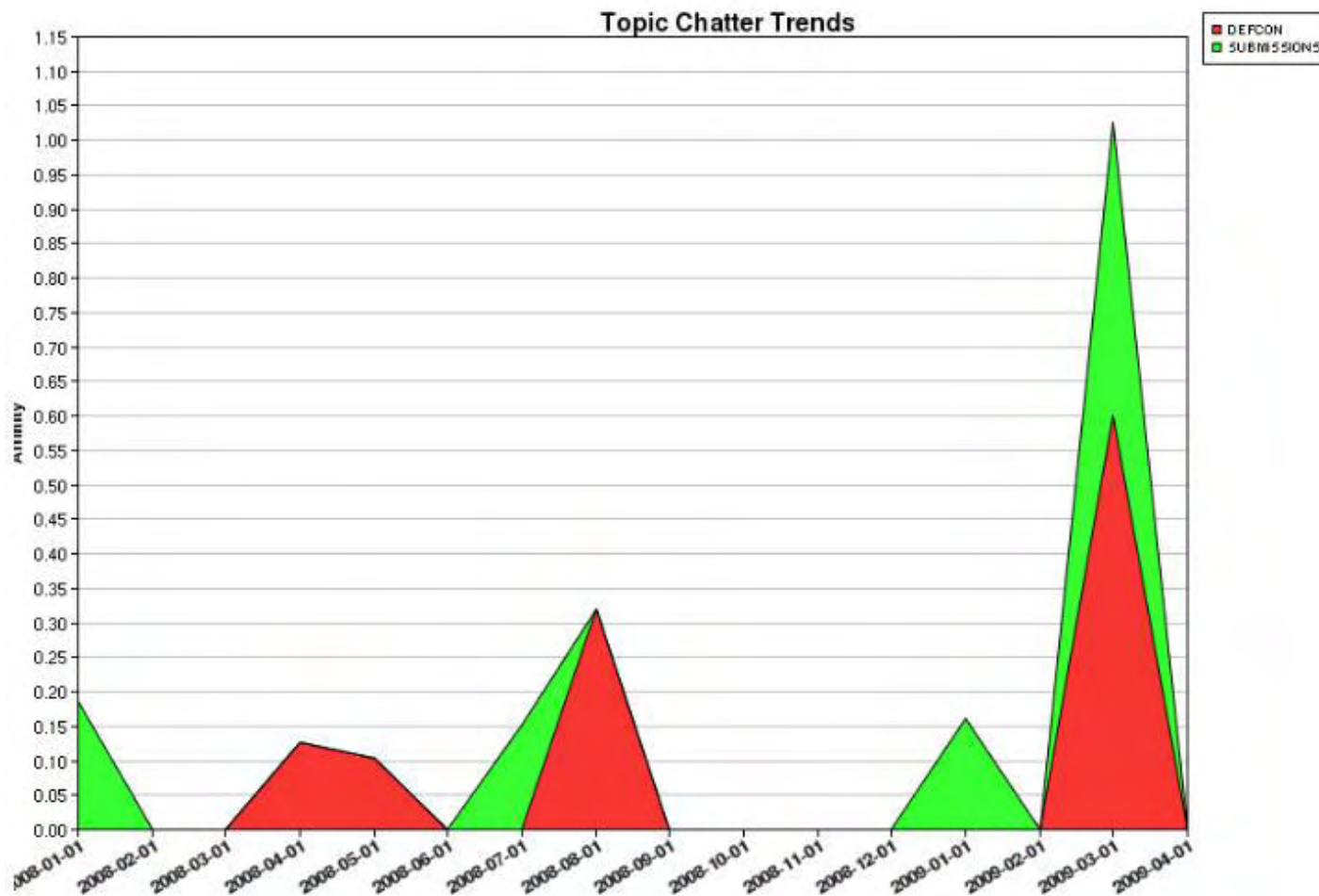
Thought Group 1

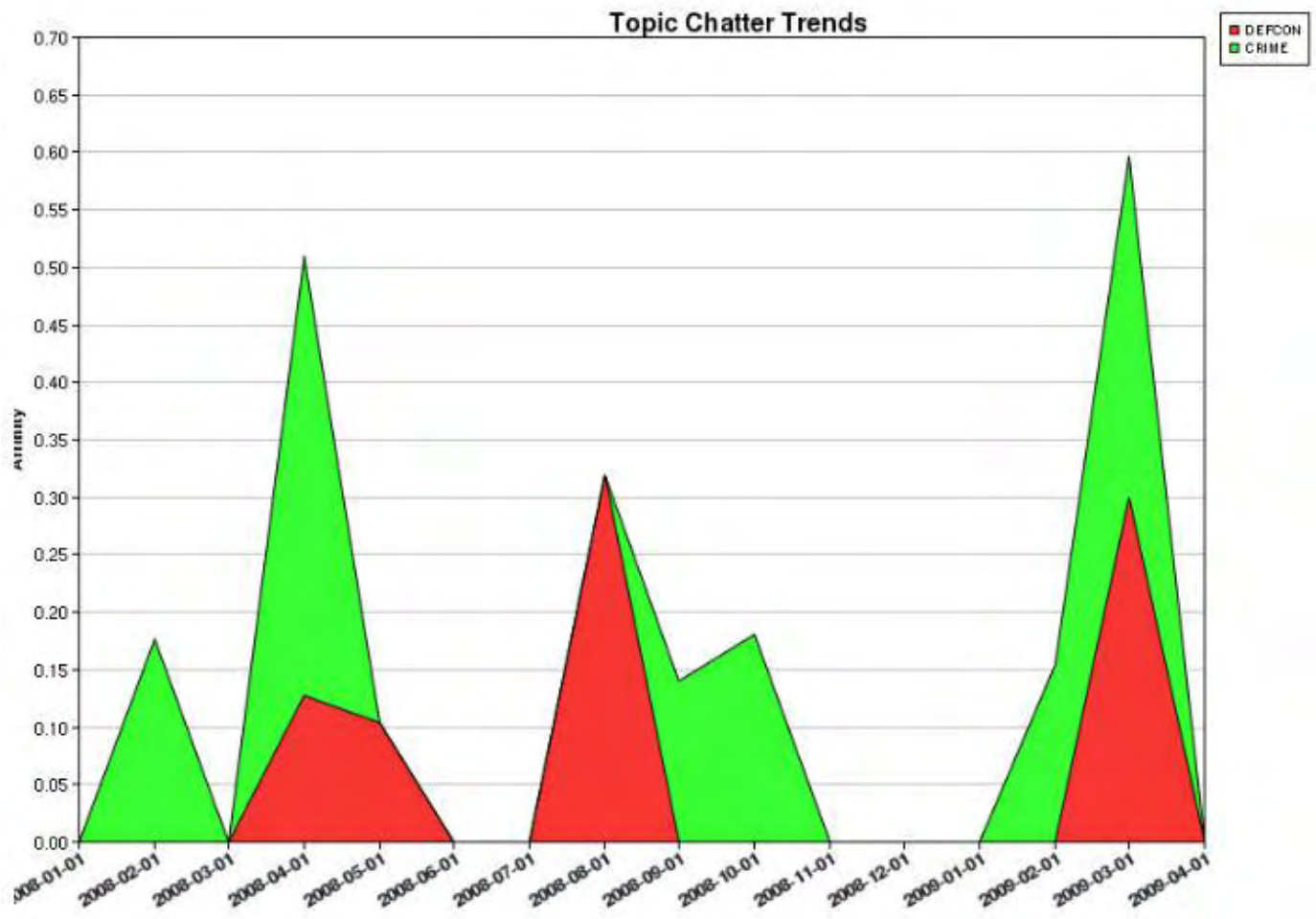
'ATTACKS' 'CHINA' 'CHINESE' 'CYBER' 'DEBIAN' 'DISCOUNT'
'ENERGY' 'FLAW' 'KEYS' 'LAB' 'LONG' 'NETWORKS'
'NUCLEAR' 'POWER' 'TAPES'

NUCLEAR	0.297	<input type="checkbox"/>
KEYS	0.225	<input type="checkbox"/>
CHINESE	0.201	<input type="checkbox"/>
POWER	0.176	<input type="checkbox"/>
CHINA	0.174	<input type="checkbox"/>
DEBIAN	0.152	<input type="checkbox"/>
DISCOUNT	0.141	<input type="checkbox"/>
ENERGY	0.139	<input type="checkbox"/>
NETWORKS	0.128	<input type="checkbox"/>
LONG	0.122	<input type="checkbox"/>
LAB	0.119	<input type="checkbox"/>
TAPES	0.117	<input type="checkbox"/>
CYBER	0.117	<input type="checkbox"/>
ATTACKS	0.116	<input type="checkbox"/>
FLAW	0.115	<input type="checkbox"/>



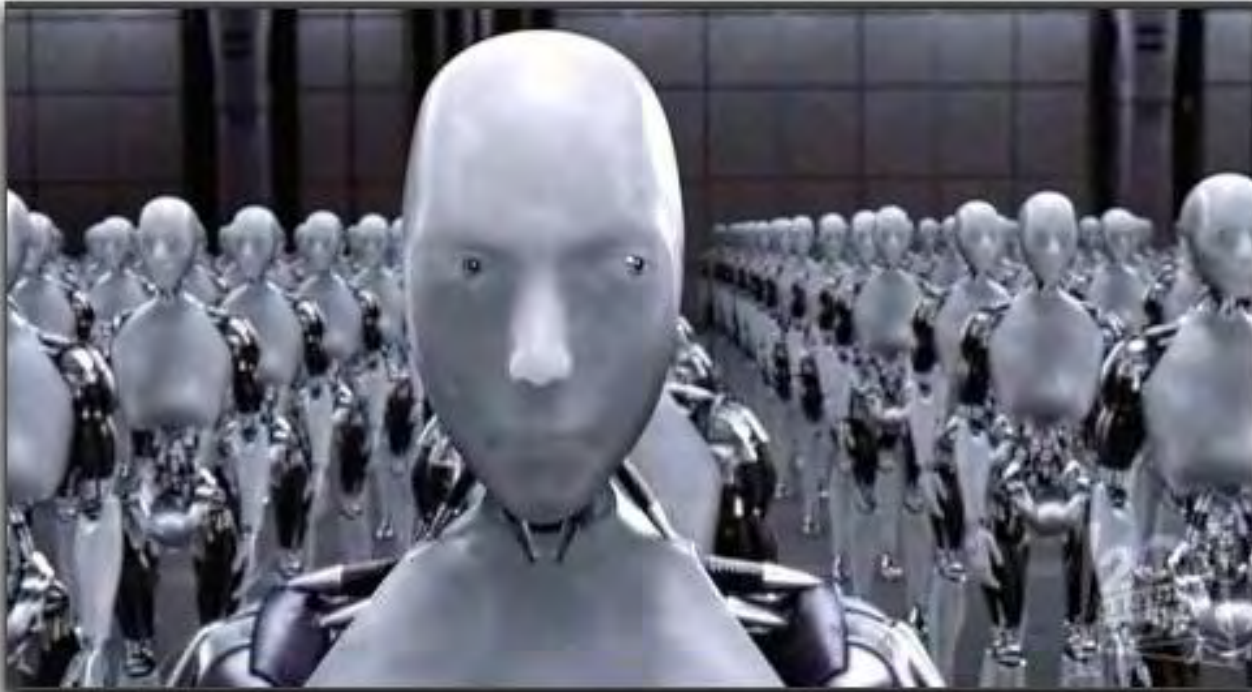






- **Defcon = Crime**







- **Grouping similar malwares together**
- **Framework**
 - **Data Collection**
 - Notes from Malware Analysts
 - **Data Storage**
 - Flat files
 - **Analysis**
 - Ontogen
 - **Decision Making**
 - Depends



Thousands of malware descriptions



Name	Date modified	Type	Size	Tags
_150	5120	agobot_fo	angarsk	appdisabler_a
_1099	7808	agobot_p	angeli	appdisabler_b
_1689	16580	agobot_q	angelina	appdisabler_c
_3008	acanze_a	agobot_vs	angus	appdisabler_d
3apa3a	achis	ai	anis	appdisabler_e
5lo	activex	aids2	anisr1	appdisabler_f
7son	ada	aih	anito_a	appdisabler_g
8-tunes	adodb_stream	aircop	anker_a	appdisabler_h
10b7	adolff	akuku	anna	appdisabler_h_dr
10past3	adore	akuma	anset	appdisabler_i
66a	adri	alabama	anthrax	appdisabler_j
99	adwdrop	alameda	antibtc	appdisabler_k
200	afcore_q	albania	anticad	appdisabler_l
217	afeto	alcarys	anticmos	appdisabler_m
268plus	afgana	aleja	antiexe	appdisabler_n
337	afpinfo	aler	antimanb	appdisabler_o
386spart	agent	alex	antimarc	appdisabler_p
403	agent007	alexande	antimon	appdisabler_q
472	agent_a	alfons	antinny	appdisabler_r
483	agent_aa	alien	antisoci	appdisabler_s
492	agent_agw	aliz	antisr1	apr1-com
555	agent_aum	aliz_dis	anto	apr1-exe
757	agent_bao	allaple_a	aol	aprilj
864	agent_bky	aloap	aolbuddy	arab
1024prsc	agent_ce	alphabet	aolcool	aragon
1308	agent_cj	alphastr	ap	arbeit
1355	agent_eo	always	ap2	archivo
1600	agiplan	ambulanc	apher	argh
1876	agobot	amoeba	aplore	armagedo
2330	agobot_ax	amus	apocalip	armagid
4870	agobot_f	andry	appder	arriba



OntoGen -- Text Garden

File Tools About

Concepts

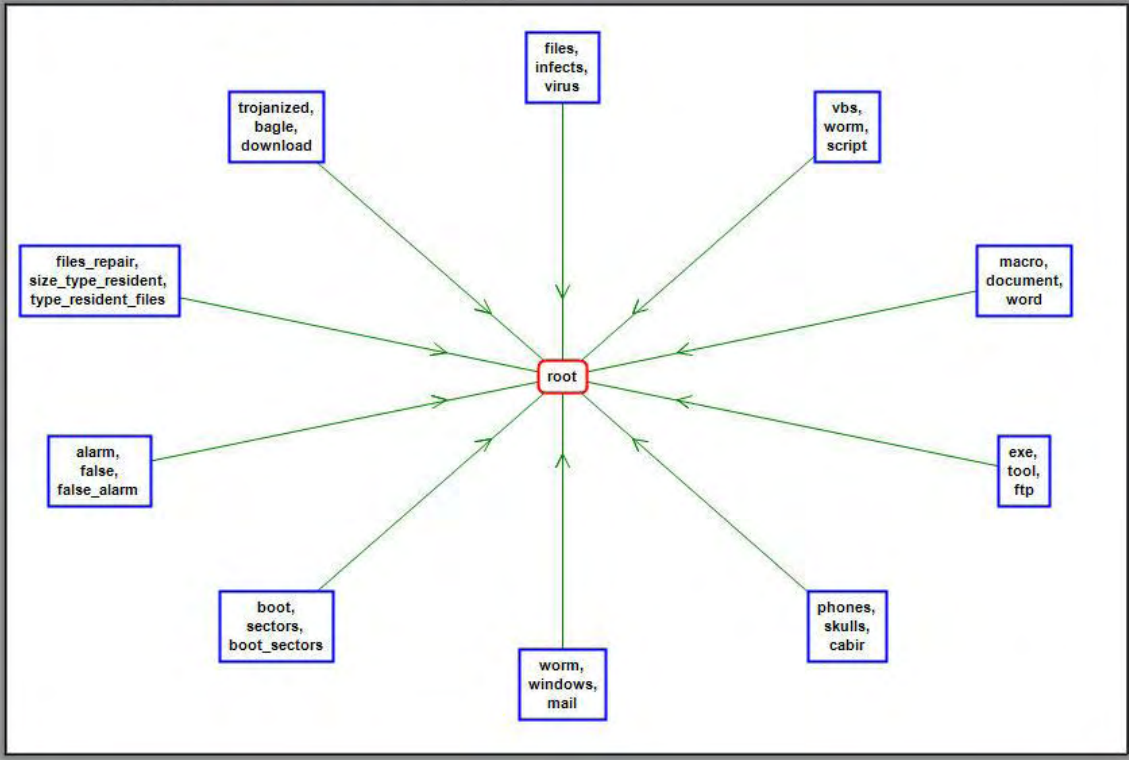
New Move Delete

- root
 - exe, tool, ftp
 - phones, skulls, cabir
 - worm, windows, mail
 - boot, sectors, boot_sectors
 - alarm, false, false_alarm
 - files_repair, size_type_resident, type_resident_files
 - trojanized, bagle, download
 - files, infects, virus
 - vbs, worm, script
 - macro, document, word

Ontology details

Ontology visualization | Concept's documents | Concept Visualization

Concept font size: Show relation type: Relation font size:



```

graph TD
    root((root))
    node1[files, infects, virus] --> root
    node2[vbs, worm, script] --> root
    node3[macro, document, word] --> root
    node4[exe, tool, ftp] --> root
    node5[phones, skulls, cabir] --> root
    node6[worm, windows, mail] --> root
    node7[boot, sectors, boot_sectors] --> root
    node8[alarm, false, false_alarm] --> root
    node9[files_repair, size_type_resident, type_resident_files] --> root
    node10[trojanized, bagle, download] --> root
    
```

Concept properties

Details | Suggestions | Relations

k-Means ▾ Suggest Query | Add Replace Prune

No. suggestions: Docs: All Unused

Keywords	No. docs	[%]

OntoGen news:



- **Monitoring of chat logs and finding “persons of interest” and who they talk to. “Cells”.**



```
<temas> but seriously I want to learn
*MrNfactor* sorry, but I get mouthy when I'm really pissed off.
<BladerHater> actually its just /var/log/messages
*** Signoff: bronc (Quit: Hey! Where'd my controlling terminal go?)

<temas> alot of people want to learn

<Debris> eCh0: know what im most curious about, for anticode.com who the hell is going
to send JP 0day

<bitshift> if I was op I wouldn't put up with this shit, ops have power for a reason

<temas> could we please let it return to that

<Coolio> BladerHater: actually its /var/adm/messages

<BladerHater> tcp wrappers is not set up on kaon
<the_Noid> yo root boyz, you need to lessen the amount of lame guest account online
<eCh0> Debris, hah..i'll send JP all my 0 day..i dont even give a shit no more.
<hook> eCh0: Ok your right, you right, we all know how right you are, now can we
please drop it?
<the_Noid> its slow as poop
<Debris> lol.
<BladerHater> no coolio actually its not
<eCh0> hook: Yes.
*** bronc (~bronc@2600.COM) has joined channel #koan
*** Mode change "+v bronc" on channel #koan by MrNfactor
```



Summary Cloud

'CHEMIST' 'CLUES' 'CONNECTING' 'DIODE' 'ANTIONLINE' 'COMMUNITY' 'GOAT' 'HATE'
'ILL' 'JERICHO' 'MCINTYRE' 'OTTER' 'ROOT' 'HTML' 'LIST' 'MACKI' 'MADE' 'MAX' 'NEAT' 'OTTER'
'SERVER' 'SHELL' 'WOW' 'PROFILE' 'TOPIC' 'USERS'

'BLADERHATER' 'BRONC' 'CHEMIST' 'CON' 'DAY' 'CHANGE' 'CONNECTION' 'MODE' 'MRNFECTOR'
'DOS' 'FRIENDS' 'IMAC' 'MODE' 'NETSHADOW' 'PAGE' 'PEER'

'PUNCHDOWN' 'SEND' 'SUPPORT' 'TEMAS'

'APR' 'BYE' 'COREY' 'GOAT' 'LINK' 'MAIL' 'BASHING' 'CODE' 'DOS' 'LINUX' 'PEOPLE'
'NEWPORTNEWS' 'TS8'

'APR' 'BITCH' 'DIALUP' 'EH' 'MAIL' 'PORT' 'ANTIONLINE' 'COOLIO' 'DENIAL'
'MOSTHATED' 'SITE' 'ZETESIS'

'BOX' 'CONNECT' 'COOLIO' 'DIODE' 'HATE' 'CODE' 'HELPING' 'LEFT' 'MACKI'
'JOINED' 'RUN' 'RUNLEVEL6' 'QUESTION' 'REPRESENT' 'STUPID'
'WARGAME'



- **Using data mining to profile companies to determine strategy**





- **Obama Town Hall Meeting**
- **Data mining of over 100,000 questions to get a “pulse” of what people are concerned about**
 - **Healthcare**



American Minds

Understanding the American Consciousness through Data Mining

Auto Industry Budget Education Financial Stability Green Jobs Energy Health Care Reform Home Ownership Jobs Retirement Security Small Business Veterans About/FAQ

'AFFORD' 'BILLS' 'DYING' 'FIELD' 'IDEAS' 'INSURED'
'LOBBYISTS' 'LONGER' 'OFFER' 'PEOPLE' 'PROGRAM' 'SICK'
'STATE' 'TYPE' 'WORKING'

'BIG' 'CHILDREN' 'DENTAL' 'ENCOURAGE' 'EXPENSES' 'FAMILIES' 'HUSBAND' 'INS'
'LOST' 'MENTAL' 'PRESCRIPTION' 'QUALIFY' 'STAND'
'UNEMPLOYMENT' 'WASHINGTON'

'AIG' 'ALTERNATIVE' 'BONUSES' 'COMMON' 'GIVING'
'HEALTHCARE' 'INCREASES' 'MAN' 'MIDDLE' 'MOVE' 'NATION'
'OWNER' 'PRACTITIONERS' 'PRIMARY' 'STATUS'

'APPROACH' 'CANNABIS' 'COMPLETE' 'EFFICIENT' 'FULL' 'THR' 'LEADING' 'OPTION'
'PAYER' 'PREVENTION' 'SINGLE' 'TABLE' 'UNIVERSAL' 'UPS' 'WONT'



'BENEFITS' 'BIRTH' 'CHARGE' 'CITIZENS' 'DONT'
'DR' 'FDA' 'FOOD' 'HOSPITALS' 'HOUSE' 'NURSE'
'POLICIES' 'PROBLEM' 'REGULATE' 'STAY'

'ASSISTANCE' 'AVERAGE' 'CHOOSE' 'CLASS' 'CONGRESS'
'COUNTRIES' 'DOCTOR' 'DOESNT' 'EXPENSIVE' 'INCOME' 'MARIJUANA'
'MEDICATIONS' 'MEDS' 'MIDDLE' 'PROFIT'

'ADS' 'BURDEN' 'CREATE' 'EMPLOYED' 'HOPE'
'ILLNESS' 'IMPROVE' 'MALPRACTICE' 'MARIJUANA'
'NURSING' 'PREVENTIVE' 'RELIEF' 'RIGHTS' 'SEX'
'WELLNESS'

'ABUSE' 'ANSWER' 'ELDERLY' 'ENGLAND' 'FAMILY' 'LAWS' 'LEGAL'
'LEGALIZING' 'MEDICAID' 'MEDICARE' 'MEDICATIONS' 'PART'
'PAYS' 'POLICY' 'PUSH'

'COST' 'COUPLES' 'DELIVERY' 'DENY' 'FAIR' 'FEDERAL' 'GOAL'
'HOSPITAL' 'INDUSTRY' 'LIVE' 'POWERFUL' 'REAL'
'SECURITY' 'SOCIAL' 'SUFFER'

'AVOID' 'CANADA' 'CANCER' 'CITIZEN' 'COVER' 'EUROPE'
'FUTURE' 'HAPPEN' 'PLACE' 'PROFITS' 'PROGRAMS' 'PUBLIC'
'STATES' 'SYSTEMS' 'UNITED'





- **Contributors**
- **Sentiment Analysis**
 - **Good or Bad?**
- **We need more data!**



- **Howard Van de Vaarst**
- **Chris Potter**
- **Secure DNA management**
- **University of Santo Tomas (Philippines)**
- **Blaz Fortuna (Ontogen)**
- **Jozef Stefan Institute, Slovenia (Text Garden)**



- **The Veritas Project**

