Primary Key Associates Limited

Scenario Forensics

Using the Primary Key Insight Engine to find new intelligence in your data

Quickly uncovering new intelligence in your digital forensic case library

Organisations that conduct digital forensic investigations can attempt to find commonality between historical cases and new investigations but often struggle with a combination of factors that impede their progress and reduce their chances of success.

- Forensic investigation resources are frequently stretched, even when working only on new cases, and have no little or no time to consider outcomes from previous cases;
- New investigators will not know all that previous investigators found, and will therefore miss potential cross-case links
- Previous searches and inferences are frequently the only information retained from an old case and may not be relevant to a new one, increasing the chance of missing key elements of commonality and new intelligence;
- Without knowledge of the *unknown knowns* in the new case data that link it with previous cases, forensic analysts will not know the searches they should be undertaking that will reveal the new evidence and intelligence.

Unknown knowns?

unknown knowns are facts and relationships that are in your data (and to that extent known) but of which you are unaware (and are therefore unknown). Unknown knowns are frequently important, but are hard to describe and expensive to find.

Scenario Forensics works without the need for initial analyst involvement

Scenario Forensics processes case data without the need for forensic analysts to have undertaken any investigation work in advance. We export case data from standard forensic tools and then operate on that data independently of any other investigation work that you may undertake. Our intelligence results improve the efficiency of both analysts and forensic investigators. Because we work on a copy of data from your standard forensic tool, it is impossible for us to accidentally corrupt or degrade the evidential quality of the data in your repository.

A parallel approach

Parallel approach avoids resource bottlenecks

By operating in parallel with human analyst driven forensic investigations we are able to enhance those investigations by uncovering the *unknown known facts* early and without creating a critical resource bottleneck.

Four stages

Scenario Forensics operates in four main stages:

- Export the new case data from existing standard forensics toolsets;
- Extract entities and relationships *from within* the new case;
- Merge the resultant output with other cases;
- Perform Scenario Analytics on merged cases.

Exporting

We use a standard scripting approach to export data from your existing forensic toolset. We have a library of our own scripts for Nuix and can develop further scripts for other tools that support this approach.

Extraction

The extraction stage, which extracts things (entities) and their relationships is technically known as "entity extraction" combined with "named entity resolution".

Entity Extraction

In Scenario Forensics *Entity Extraction* uses a variety of approaches to scavenge potential entities from files. These do not have to be of a particular type, since by its nature digital forensics has to deal with unknown file types, and interesting information can exist in a variety of places. In principle, this approach allows Scenario Forensics to use information in database files without knowing the schema, as can easily happen when dealing with partial acquisition.

Potential entities and relationships can be recognised by their morphological structure – i.e. what they look like. However, this can easily be in error, so we use three techniques to improve the results:

- We use further rules, with additional knowledge, to reduce the candidate set of entities and relationships.
- We infer some links without needing specific textual link evidence.
- Finally we are content to allow this stage to slightly over-retrieve (which is better than under-retrieving) since invalid associations will normally be ignored in later stages, which are bringing further domain knowledge to bear.

Named Entity Resolution

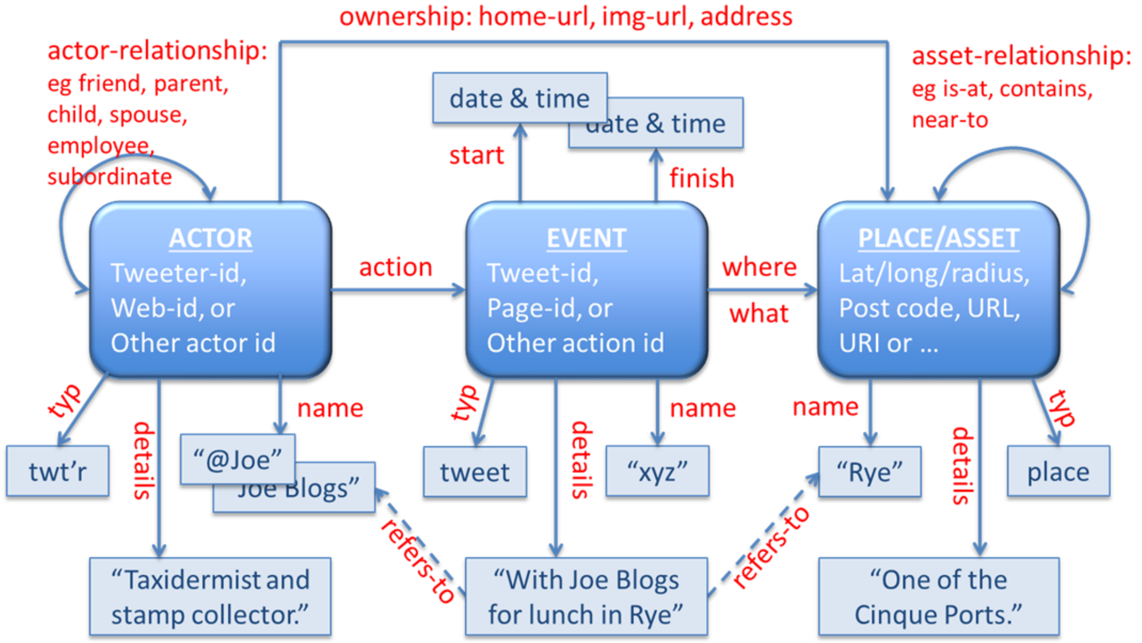
Named Entity Resolution is the problem of working out which entity name goes with which entity. For example, there could easily be more than one "John Brown", so it is hard for the system to know how many there are, and which "John Brown" is which, as well as whether "J Brown" is "John Brown" or indeed someone else entirely. Again, we use the fact that Scenario Analytics has domain knowledge to apply later on, and therefore we do not have to fully resolve entities at this stage: we are better off deferring this problem to the Scenario Analytics. From an information perspective, this makes sense: a human analyst might also not be able to resolve "John Brown" until he too had further information.

Merge with other cases

Individual cases are subsequently *merged* with other cases, so that cross-case comparison can occur. A major advantage of this approach is that when an individual case must be deleted (perhaps for legal reasons), that case can be deleted, and the remainder re-merged, rather than having to re-export and re-extract all the cases.

The Insight Engine has a built-in social model or schema

The Insight Engine's schema models the social world in which people and organisations exist in places, do things at points in time, communicate, and have various types of relationships. The model is intrinsically flexible, and adds the concept of evidence to the Primary Key Three Entity model shown below. This is very helpful in explaining *why* the Insight Engine considers there to be a link between two entities. The analyst can assess the strength of that evidence, chase down resultant leads using the uncovered links, or choose to find the original document (for example) in the forensics tool for inclusion in a case report.


ownership: home-url, img-url, address
actor-relationship:
eg friend, parent, child, spouse, employee, subordinate
asset-relationship:
eg is-at, contains, near-to
date & time
date & time
start
finish
ACTOR
Tweeter-id, Web-id, or Other actor id
action
EVENT
Tweet-id, Page-id, or Other action id
where
what
PLACE/ASSET
Lat/long/radius, Post code, URL, URI or …
typ
details
name
typ
details
name
name
details
typ
twt'r
"@Joe"
"Joe Blogs"
tweet
"xyz"
"Rye"
place
"Taxidermist and stamp collector."
refers-to
"With Joe Blogs for lunch in Rye"
refers-to
"One of the Cinque Ports."

Analyse with Scenario Analytics

The core of the Insight Engine is powered by Scenario Analytics, which is based around the concept of a *scenario* - a combination of circumstances that an investigator would find interesting. For example, an investigator will be interested in any indicators that a person of interest is using aliases, and what those aliases may be. (In this example, the other aliases are unknown-knowns: they are 'known' in the data, but without the Insight Engine, remain unknown to your organisation.)

Scenarios built in to the Insight Engine

The Insight Engine comes with a *built-in suite* of such scenarios, many of which can be customised on the fly, and some of which only make sense when customised, such as "Who has contact with person X". As is normal with Scenario Analytics, the solutions to these scenarios are explained in natural language, and can be shown as a helpful diagram, explaining not just the "what" (i.e. who has contact) but how and why Scenario Analytics had come to that conclusion.

Adding new scenarios

We can add new scenarios, with each one taking about one day of effort, including both the time of the requesting analyst and the knowledge engineer who builds the scenario in the SDL ("Scenario Discovery Language") used behind the scenes. (For those interested, SDL is a declarative language: you write what you want to find, not how to find it, which is sorted out by the applied machine intelligence within Scenario Analytics.) By building new scenarios, expert knowledge of an experienced investigator can be made available to junior analysts, since the SDL scenario has captured the expert's insight.

Adding information by hand

The analyst can also add information into the Insight Engine 'by hand'. So if it is already known that Mr A sells stolen goods to Mrs B, or Dr C is an alias of Miss D, then such facts can be directly added to the cases, and the Insight Engine can then make use of them.

About us

Since 2010 Primary Key Associates has delivered world-class consultancy, developing and exploiting cutting-edge analytics, artificial intelligence and digital investigation technologies.

Answer-Focused Technologies

We re-invest much of our revenue to develop cutting-edge data analytics, investigation and artificial intelligence technology to address the business problems we see our clients facing. Our IPR portfolio includes:

- Primary Key **Scenario Analytics** – a fast, modern and flexible data analytics technology to find, explain and illustrate the 'unknown knowns' in your large datasets.
- Primary Key **Insight Engine** – Scenario Analytics combined with powerful entity and relationship extraction, applied to digital forensics that reveals previously unknown connections in evidence.
- Primary Key **Illuminate** – a social media analysis and intelligence technology and service to find and analyse open source data to address particular business problems.
- Primary Key **Incipients** – predictive analytics that identify what is going to trend.
- Primary Key **Distil** – a technology toolkit we deploy on client sites to both find frauds and identify the business practices that make you vulnerable to fraud.
- **Fast analytic heuristics** to overcome the bottleneck of solving intrinsically non-parallel, yet vital, problems.

Expert consultancy and services

Our team are experienced professionals in IT and related areas:

- As experts in enterprise and security architectures (including SABSA), information security and cryptography we help design, validate, and test systems which are secure both physically and in cyber-space.
- We build software systems from the smallest (in C and assembler coded microcontrollers) through the mobile (Java or Swift coded Android or IoS apps), to the ubiquitous (Python or C+ on PCs, web technologies and virtualised servers) and even the esoteric (real-time and spacecraft).
- We understand data (SQL and graph databases) and artificial intelligence (computational linguistics, image analysis and machine learning) and have fast algorithms for timely and secure analysis of big datasets and evidential analysis of social media.
- We provide cyber threat intelligence and competitive intelligence.
- We support civil and criminal digital investigations from open source research through digital forensic analysis and providing expert witness reports and testimony.
- We undertake both technical and business programme and project management.

e: feedback@primarykey.co.uk
w: www.primarykey.co.uk
t: +44 1403 599900

@pkaluk

SWEET & MAXWELL
CHECKED
2017
EXPERT WITNESS