Investigating People: A Qualitative Analysis of the Search Behaviours of Open-Source Intelligence Analysts

Leif Azzopardi
University of Glasgow
Leif.Azzopardi@glasgow.ac.uk

Sean McKeown
University of Glasgow
s.mckeown.1@research.gla.ac.uk

David Maxwell
University of Glasgow
d.maxwell.1@research.gla.ac.uk

Wm. Bradley Glisson
University of South Alabama
bglisson@southalabama.edu
Open-Source INTelligence (OSINT)

• *Intelligence* created from **public data** (open-source)

• Source Examples:
  – Web
  – News media
  – Google maps

• Many uses cases for OSINT
  – Country level, business level, personal level
OSINT Use Cases

- Focus on personal level
  - Insurance Fraud
  - Employee Vetting
  - Law Enforcement
OSINT Use Cases

- Focus on personal level

  - **Insurance Fraud**  
    Insurance Fraud Bureau - 2014  
    Costs the UK economy **£2.1 Billion** annually

- Employee Vetting

- Law Enforcement
OSINT Use Cases

- Focus on personal level
  - Insurance Fraud  
    Costs the UK economy **£2.1 Billion** annually  
    Insurance Fraud Bureau - 2014
  - Employee Vetting  
    43% of employers vet applicants using social media  
    Careerbuilder - 2014
  - Law Enforcement
OSINT Use Cases

• Focus on personal level
  
  – Insurance Fraud  
  Insurance Fraud Bureau - 2014
  Costs the UK economy **£2.1 Billion** annually

  – Employee Vetting  
  Careerbuilder - 2014
  **43% of employers** vet applicants using social media

  – Law Enforcement  
  Lexis Nexis - 2012
  **4 out of 5** law enforcement officers in the U.S.A. use social media for investigative purposes
EXPLORING SEARCH BEHAVIOUR
Analyst Search Behaviour

- Want to find out about how the analysts find/search for information, analyse data
  - Existing literature only covers high level processes
## Analyst Search Behaviour

- Want to find out about how the analysts find/search for information, analyse data
  - Existing literature only covers high level processes

- Analogous work for closed intelligence analysis
  - Chin et al – 2009
  - Pirolli and Card – 2005
Analyst Search Behaviour

• Want to find out about how the analysts **find/search for information, analyse data**
  – Existing literature only covers high level processes

• Analogous work for closed intelligence analysis
  – Chin et al – 2009
  – Pirolli and Card – 2005

• Frame as an instance of **domain specific search**.
  – Such as: *Web search, Patent search, E-Discovery*
  – Commonalities, shared properties?
Qualitative Study

• Worked with a leading investigation company
• Small scale, exploratory, case study.
  – Initial probe of user behaviour.
Qualitative Study

- Worked with a leading *investigation company*
- Small scale, *exploratory*, case study.
  - Initial probe of user behaviour.

**Executive interview** *(Context, expectations, groundwork)*
- Overall business practices
- Types of investigation undertaken, tasks
- Executive view of analyst processes, behaviour, report format
- Used to construct *questionnaire*
Qualitative Study

• Worked with a leading *investigation company*
• Small scale, *exploratory*, case study.
  – Initial probe of user behaviour.

Executive interview (*Context, expectations, groundwork*)
• Overall business practices
• Types of investigation undertaken, tasks
• Executive view of analyst processes, behaviour, report format
• Used to construct *questionnaire*

Semi-Structured Interviews with two analysts (*Behaviour*)
• *Questionnaire* used as guide posts for interviews
• Investigative *processes, search behaviours, difficulties*, anecdotes.
• Tools, services
FINDINGS
Overall Investigative Process

Client → Briefing Sheet

Report of Findings

Analyst

Updated Briefing Sheet

Surveillance Operative

Social Network

WWW

Subject
Overall Investigative Process

Client → Briefing Sheet → Analyst → Updated Briefing Sheet → Surveillance Operative

Report of Findings

Social Network

WWW

Subject
Overall Investigative Process

Client > Briefing Sheet > Analyst > Report of Findings > Surveillance Operative

Updated Briefing Sheet

Social Network

WWW

Subject
Overall Investigative Process

Client

Briefing Sheet

Report of Findings

Analyst

Updated Briefing Sheet

Surveillance Operative

Social Network

WWW

Subject
Advancing the Investigation

• Small pieces of information
  – The operating unit of the investigation
  – Expand / narrow the investigation
Advancing the Investigation

• Small pieces of information
  – The operating unit of the investigation
  – Expand / narrow the investigation

• Example information types
  – “The basics” – Name, address, D.O.B, etc.
Advancing the Investigation

• Small pieces of information
  – The operating unit of the investigation
  – Expand / narrow the investigation

• Example information types
  – “The basics” – Name, address, D.O.B, etc.
  – Hobbies, activities, associations, memberships, ownerships
    • e.g. Plays darts, supports a football club, runs marathons, owns a Ferrari
Advancing the Investigation

• Small pieces of information
  – The **operating unit of the investigation**
  – Expand / narrow the investigation

• Example information types
  – “The basics” – Name, address, D.O.B, etc.
  – Hobbies, activities, associations, memberships, ownerships
    • e.g. Plays darts, supports a football club, runs marathons, owns a Ferrari
  – Contextual information
    • e.g. Metadata, history, politics, locale, geospatial relations
Investigator’s Briefing Sheet

Subject: John Smith
Marital Status: Married
Occupation: Unemployed. Previously employed as a supermarket manager.
Address: 26 Oldfield Ave, London.

Benefits: Mr. Smith is claiming disability allowance due to an injury incurred in an accident.

Accident: Mr. Smith was involved in a traffic accident while riding his bicycle to his previous place of employment.

Injury: Mr. Smith claims that he is unable to stand up and cannot take care of himself.

Sergei Scholtz
Branch Manager

Financo Ltd.
23 New Street
G99 3RT
Investigator’s Briefing Sheet

Subject: John Smith
Marital Status: Married
Occupation: Unemployed. Previously employed as a supermarket manager.
Address: 26 Oldfield Ave, London.
Benefits: Mr. Smith is claiming disability allowance due to an injury incurred in an accident.
Accident: Mr. Smith was involved in a traffic accident while riding his bicycle to his previous place of employment.
Injury: Mr. Smith claims that he is unable to stand up and cannot take care of himself.

Sergei Scholtz
Branch Manager
Benefits Example - Cyclist

Verify Briefing Sheet
Verify Briefing Sheet

- Note circumstances of accident – **cycling**
Benefits Example - Cyclist

Verify Briefing Sheet

• Note circumstances of accident – cycling
• No longer working at previous employer
Verify Briefing Sheet

- Note circumstances of accident – cycling
- No longer working at previous employer
- Residence erroneous
  - Moved residence from London to Newcastle
Verify Briefing Sheet

- Note circumstances of accident – **cycling**
- No longer working at previous employer
- Residence erroneous
  - **Moved residence** from London to **Newcastle**

Web Search
Benefits Example - Cyclist

Verify Briefing Sheet
• Note circumstances of accident – cycling
• No longer working at previous employer
• Residence erroneous
  – Moved residence from London to Newcastle

Web Search
• Facebook profile is completely private
Verify Briefing Sheet
- Note circumstances of accident – **cycling**
- No longer working at previous employer
- Residence erroneous
  - **Moved residence** from London to **Newcastle**

Web Search
- Facebook profile is completely private
  - Co-resident’s profile isn’t, exposes more social relations/family
  - Find many **comments about a race**
Benefits Example - Cyclist

Verify Briefing Sheet
• Note circumstances of accident – cycling
• No longer working at previous employer
• Residence erroneous
  – Moved residence from London to Newcastle

Web Search
• Facebook profile is completely private
  – Co-resident’s profile isn’t, exposes more social relations/family
  – Find many comments about a race
• Check for events in the Newcastle area via Google
Verify Briefing Sheet

- Note circumstances of accident – **cycling**
- No longer working at previous employer
- Residence erroneous
  - Moved residence from London to **Newcastle**

Web Search

- Facebook profile is completely private
  - Co-resident’s profile isn’t, exposes more social relations/family
  - Find many **comments about a race**
- Check for events in the Newcastle area via Google
  - Find the subject listed as a **participant in a recent cycling race** (post claim date)
Background Verification Stage

- Relatively Fixed Services
  - May change with investigation type

Electoral Roll

Property

Corporate Directory

• Relatively Fixed Services
  – May change with investigation type

Briefing Sheet

Analyst

Updated Briefing Sheet
Background Verification Stage

- Relatively Fixed Services
  - May change with investigation type
- Service overlap for corroboration

**Diagram:**
- Compare results
- Briefing Sheet
- Analyst
- Updated Briefing Sheet

**Analysis:**
- Relatively Fixed Services
  - May change with investigation type
- Service overlap for corroboration
Background Verification Stage

- Relatively Fixed Services
  - May change with investigation type
- Service overlap for corroboration
  - Same query submitted to multiple services

- “John Smith”

Briefing Sheet → Analyst → Updated Briefing Sheet
Background Verification Stage

- Relatively Fixed Services
  - May change with investigation type
- Service overlap for corroboration
  - Same query submitted to multiple services
- Preference for unpaid services
  - Micro-transactional process
Open-Ended Stage

- Emphasis on Social Media
  - Lateral utility, friends and relatives profiles

![Diagram showing analyst, search engine, social media, and wider web connections]
Open-Ended Stage

• Emphasis on Social Media
  – Lateral utility, friends and relatives profiles
• Still a small number of fixed sources
  – e.g. Local news sites, local court information.
Open-Ended Stage

• Emphasis on Social Media
  – Lateral utility, friends and relatives profiles
• Still a small number of fixed sources
  – e.g. Local news sites, local court information.
• Search Engine as a gateway to the wider web
  – New information from previously unknown sources.
# Analyst Search Behaviour

<table>
<thead>
<tr>
<th></th>
<th>Analyst 1</th>
<th>Analyst 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Analyst Search Behaviour

<table>
<thead>
<tr>
<th>Search Strategy</th>
<th>Analyst 1</th>
<th>Analyst 2</th>
</tr>
</thead>
</table>
| **Query orientated** | - Experimental Queries  
- Discriminatory when examining documents  
- Max 8 hours, may return to search later during periods of downtime | **Document orientated**  
- Examines many documents per query  
- Initially 15 pages per query, more if required  
- Potentially hundreds of results, tens of pages |
## Analyst Search Behaviour

<table>
<thead>
<tr>
<th></th>
<th>Analyst 1</th>
<th>Analyst 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search Strategy</td>
<td><strong>Query orientated</strong></td>
<td><strong>Document orientated</strong></td>
</tr>
<tr>
<td></td>
<td>• Experimental Queries</td>
<td>• Examines many documents per query</td>
</tr>
<tr>
<td></td>
<td>• Discriminatory when examining documents</td>
<td>• Initially 15 pages per query, more if required</td>
</tr>
<tr>
<td></td>
<td>• Max 8 hours, may return to search later during periods of downtime</td>
<td>• Potentially hundreds of results, tens of pages</td>
</tr>
<tr>
<td>Advanced Features</td>
<td>• None used</td>
<td>• Boolean operator</td>
</tr>
<tr>
<td></td>
<td>- self attributed to lack of skill</td>
<td>• Phrase queries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reverse image searching</td>
</tr>
</tbody>
</table>
## Analyst Search Behaviour

<table>
<thead>
<tr>
<th></th>
<th>Analyst 1</th>
<th>Analyst 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search Strategy</strong></td>
<td><strong>Query orientated</strong></td>
<td><strong>Document orientated</strong></td>
</tr>
<tr>
<td></td>
<td>• Experimental Queries</td>
<td>• Examines many documents per query</td>
</tr>
<tr>
<td></td>
<td>• Discriminatory when examining documents</td>
<td>• Initially 15 pages per query, more if required</td>
</tr>
<tr>
<td></td>
<td>• Max 8 hours, may return to search later during periods of downtime</td>
<td>• Potentially hundreds of results, tens of pages</td>
</tr>
<tr>
<td><strong>Advanced Features</strong></td>
<td>• None used</td>
<td>• Boolean operator</td>
</tr>
<tr>
<td></td>
<td>- self attributed to lack of skill</td>
<td>• Phrase queries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Reverse image searching</td>
</tr>
<tr>
<td><strong>Common</strong></td>
<td>• No collaboration during web search</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• No well defined stopping point!</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Use judgment to know when to stop</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Not required to find all evidence, just sufficient evidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Use of contextual and background knowledge</td>
<td></td>
</tr>
</tbody>
</table>
# Other Analyst Behaviour

<table>
<thead>
<tr>
<th></th>
<th>Analyst 1</th>
<th>Analyst 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reliability</td>
<td>• Generally taken at face value unless suspicious</td>
<td>• Generally sceptical</td>
</tr>
<tr>
<td></td>
<td>• Wary of particular types of information</td>
<td>• Distinction between ‘facts’ and ‘reported facts’</td>
</tr>
</tbody>
</table>

- Generally taken at face value unless suspicious
- Wary of particular types of information
- Generally sceptical
- Distinction between ‘facts’ and ‘reported facts’
### Other Analyst Behaviour

<table>
<thead>
<tr>
<th></th>
<th>Analyst 1</th>
<th>Analyst 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
<td>• Generally taken at face value unless suspicious</td>
<td>• Generally sceptical</td>
</tr>
<tr>
<td></td>
<td>• Wary of particular types of information</td>
<td>• Distinction between ‘facts’ and ‘reported facts’</td>
</tr>
<tr>
<td><strong>Digital Footprints</strong></td>
<td>• Uses services requiring user accounts if known to be safe</td>
<td>• Categorically avoids services which require user accounts</td>
</tr>
</tbody>
</table>
## Other Analyst Behaviour

<table>
<thead>
<tr>
<th></th>
<th>Analyst 1</th>
<th>Analyst 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reliability</strong></td>
<td>• Generally taken at face value unless suspicious</td>
<td>• Generally sceptical</td>
</tr>
<tr>
<td></td>
<td>• Wary of particular types of information</td>
<td>• Distinction between ‘facts’ and ‘reported facts’</td>
</tr>
<tr>
<td><strong>Digital Footprints</strong></td>
<td>• Uses services requiring user accounts if known to be safe</td>
<td>• Categorically avoids services which require user accounts</td>
</tr>
<tr>
<td><strong>Reporting</strong></td>
<td>• Screenshots</td>
<td>• Descriptive notes</td>
</tr>
<tr>
<td></td>
<td>• Qualifications</td>
<td>• Qualifications</td>
</tr>
<tr>
<td></td>
<td>• URLs / Sources</td>
<td>• URLs / Sources</td>
</tr>
<tr>
<td></td>
<td>• ‘Indicators’ as opposed to ‘facts’</td>
<td>• ‘Indicators’ as opposed to ‘facts’</td>
</tr>
<tr>
<td></td>
<td>Analyst 1</td>
<td>Analyst 2</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Reliability</td>
<td>• Generally taken at face value unless suspicious</td>
<td>• Generally sceptical</td>
</tr>
<tr>
<td></td>
<td>• Wary of particular types of information</td>
<td>• Distinction between ‘facts’ and ‘reported facts’</td>
</tr>
<tr>
<td>Digital Footprints</td>
<td>• Uses services requiring user accounts if known to be safe</td>
<td>• Categorically avoids services which require user accounts</td>
</tr>
<tr>
<td>Reporting</td>
<td>• Screenshots</td>
<td>• Descriptive notes</td>
</tr>
<tr>
<td></td>
<td>• Qualifications</td>
<td>• Qualifications</td>
</tr>
<tr>
<td></td>
<td>• URLs / Sources</td>
<td>• URLs / Sources</td>
</tr>
<tr>
<td></td>
<td>• ‘Indicators’ as opposed to ‘facts’</td>
<td>• ‘Indicators’ as opposed to ‘facts’</td>
</tr>
<tr>
<td>Common</td>
<td>• Reports: PDFs and Customer Relationship Management (CRM) system</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• No conclusions or recommendations</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>No browser add-ons, meta-search tools, case management software</strong></td>
<td></td>
</tr>
</tbody>
</table>
Difficulties - Names

• Person disambiguation
  – Does this information/profile belong to the subject?
  – Uncertainty, corroboration, qualifications in report

• Name Variants
  – Misspellings, alternative spellings, colloquialisms, pet names
  – Misleading, may cause documents to be missed
Difficulties - Volume

- Stagnation
  - Caused by difficulty in finding initial pieces of information
  - Once a thread is picked up, progress follows more easily

- Large Quantities of Results (*primary concern*)
  - Web search engines produce many results
  - Takes time to trawl through results, assess relevance, extract information
  - “*Minimum time in for maximum information out*”
• Stagnation
  – Caused by difficulty in finding initial pieces of information
  – Once a thread is picked up, progress follows more easily

• Large Quantities of Results (*primary concern*)
  – Web search engines produce many results
  – Takes time to trawl through results, assess relevance, extract information
  – “Minimum time in for maximum information out”
Summary

• **Domain Expertise**, as opposed to search expertise
  – Follow a pattern, invested in results, avoid commercial sites
Summary

• **Domain Expertise**, as opposed to search expertise
  – Follow a pattern, invested in results, avoid commercial sites

• **Exploratory search**, rather than exhaustive
  – No requirement to find all pieces of evidence
Summary

- **Domain Expertise**, as opposed to search expertise
  - Follow a pattern, invested in results, avoid commercial sites

- **Exploratory search**, rather than exhaustive
  - No requirement to find all pieces of evidence

- Reports contain **no hypotheses, no conclusions**, in contrast to closed intelligence analysts
**Summary**

- **Domain Expertise**, as opposed to search expertise
  - Follow a pattern, invested in results, avoid commercial sites

- **Exploratory search**, rather than exhaustive
  - No requirement to find all pieces of evidence

- Reports contain **no hypotheses, no conclusions**, in contrast to closed intelligence analysts

- **Reliability** concerns in line with those of closed intelligence analysts
  - Directly comparable to two analysts in Chin et al 2001
Limitations and Future Work

• Small scale study
  – Do the results Generalise?
  – Expand to larger user base via survey
Limitations and Future Work

• Small scale study
  – Do the results Generalise?
  – Expand to larger user base via survey

• Person centric OSINT only
  – Differences with wider OSINT use cases?
  – Conduct similar research for related domains
Limitations and Future Work

• Small scale study
  – Do the results Generalise?
  – Expand to larger user base via survey

• Person centric OSINT only
  – Differences with wider OSINT use cases?
  – Conduct similar research for related domains

• No difficulty mitigation
  – Explore computational mechanisms to address analyst difficulties
Limitations and Future Work

• Small scale study
  – Do the results Generalise?
  – Expand to larger user base via survey

• Person centric OSINT only
  – Differences with wider OSINT use cases?
  – Conduct similar research for related domains

• No difficulty mitigation
  – Explore computational mechanisms to address analyst difficulties

• Explore the legal context
References

• **Careerbuilder** 2014  

• **Insurance Fraud Bureau** 2014  
  http://www.insurancefraudbureau.org/

• **Lexis Nexis** 2012  
  http://www.lexisnexis.com/government/investigations/

• **G. Chin,Jr., O. A. Kuchar, and K. E. Wolf.** *Exploring the analytical processes of intelligence analysts.* 2001

• **P. Pirolli and S. Card.** *The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis.* 2005

http://www.seniorliving.org/