2013

A study on the effects of tertiary education on open source information gathering skills

Sara T.K. Cunow
Edith Cowan University

2013

A study on the effects of tertiary education on open source information gathering skills

Sara T.K. Cunow
Edith Cowan University

Edith Cowan University

Copyright Warning

A Study on the Effects of Tertiary Education on Open Source Information Gathering Skills

A dissertation submitted in partial fulfilment of the requirements for the degree

Y85 Bachelor of Science (Security) Honours

By

Sara Teresa Kasprzyk Cunow

Student # 10160797

School of Computer and Security Science

Faculty of Computing, Health & Science

Edith Cowan University

Supervisors: Lecturer Mr Jeffery Corkill & Senior Lecturer Dr Eyal Gringart

Submission Date: 26st of July 2013

ABSTRACT

Since the introduction of the World Wide Web (WWW), a large amount of information has become available and accessible to anyone with an Internet connection. Where in the past, the art of the intelligence profession was finding scarce information, currently information gathering is more focused on sorting relevant information from the available abundance. The purpose of the current study was to gain a better understanding of how information is gathered on the Web by potential intelligence analysts. Although the WWW is used by many people to search for information daily, relatively little research exists on how this source should be used and what to consider when using it in the intelligence context. As the intelligence profession mostly recruits university graduates as analysts, the current study aimed to investigate how information collection skills differ between security science students who are at different stages of a three-year tertiary intelligence course.

A mixed-method approach was employed using three cohorts of students with 40 participants. Each participant was asked to gather information on a defined problem utilising resources available on the WWW, to list all information gathered and the search terms used. In addition, each participant was asked to specify search strategies employed to address the problem, which were analysed qualitatively. Statistical tests were used to determine statistically significant differences between the three levels of cohorts concerning volume of information gathered, number of search terms utilised and number of clicks used. It was found that the second year cohort utilised a statistically significant greater number of search terms than the first year cohort. Qualitative data were analysed to identify that eight strategies overall, varying in frequency of use and level of sophistication, were used by participants at different stages of the course. The greatest searching skill acquisition was found to occur in the first year of the course, followed by further refinements of skills throughout the second and third year of the course. Replication of the study is recommended and future research directions are suggested.

COPYRIGHT AND ACCESS DECLARATION

I certify that this thesis does not, to the best of my knowledge and belief:

(i) Incorporate without acknowledgment any material previously submitted for a degree or diploma in any institution of higher education;

(ii) Contain any material previously published or written by another person except where due reference is made in the text; or

(iii) Contain any defamatory material

Signature  .....

Date 26 July 2013 .....

ACKNOWLEDGEMENTS

I would like to start by expressing my eternal gratitude to Dr Eyal Gringart and Mr Jeffery Corkill, who both had supervisory roles and provided invaluable guidance in completing this thesis.

I would also like to thank Associate Professor Andrew Woodward and Mr Michael Coole for their kindness, selfless availability and individual support throughout this year. I am extremely humbled and grateful for your never-ending encouragements and belief in me.

To all those who took the time and participated in this study, your cooperation and assistance has been invaluable and I thank each of you for taking part. This thesis could not have been accomplished without you.

Finally, an extreme debt of appreciation and gratitude has to go to my partner Kane; I could not have achieved this without your selflessness, patience and understanding. And to my family and friends for providing the base that allows me to achieve all that I do. Without your humour and constant support I would not have been able to retain my sanity while completing this thesis.

DEFINITION OF TERMS

Intelligence: a product that is the result of information that has been gathered from a variety of sources, analysed and integrated and produced into a product to support decision-making.

Intelligence analyst: somebody whose primary employment role is to take information and turn it into an intelligence product to support a decision-maker.

Intelligence cycle: the process of intelligence production is commonly referred to as the intelligence cycle, including the following five steps; direction, collection, processing, analysis and dissemination.

Decision-maker: a person who selects what actions to take amongst more than one option and who operates "within time and cognitive limitations that prevent them from evaluating all possible decisions" (Agosto, 2002, p. 16).

ABBREVATIONS

Abbreviations that appear more than once in the thesis are listed below.

HUMINT	Human Intelligence
IMINT	Imagery Intelligence
MASINT	Measurement and Signatures Intelligence
OSINT	Open Source Intelligence
SIGINT	Signals Intelligence
WWW	World Wide Web, also referred to as the Web

TABLE OF CONTENTS

Appendices

LIST OF TABLES

LIST OF FIGURES

"Facts are not everything – at least half the business lies in how you interpret them."

Fyodor Dostoevsky

Chapter 1: Introduction

Introduction

Intelligence is understood to be a process that revolves around the collection of information, the analysis of that information and the production of a product that aids with decision-making. Thus, information collection is a significant part of the production of intelligence, as without information one cannot produce intelligence (Prunckun, 2010). Historically, the focus of collection was on classified and clandestine systems as information was secret and difficult to acquire. However, with recent technological advances, an abundance of information has become openly available and, making billions of pieces of information available to anyone connected to the Internet, the World Wide Web (WWW) has become the largest information store in the world (Downes, 2007). Thus, shifting the focus of intelligence information foraging from that of finding secret and scarce information to identifying what is relevant from the vast knowledge that is available through open sources (Moore, 2011; Olcott, 2012; Quarmby & Young, 2010). As expressed by the former chair of the National Intelligence Council, Joseph Nye, "open source intelligence is the outer piece of the jigsaw puzzle . . . open source intelligence is the critical foundation for the all-source intelligence product" (Cited in Sims & Gerber, 2005, p. 64).

Little information exists on how intelligence analysts go about foraging information via open sources such as the WWW, as it is relatively new and still not well understood. Hence, the current study investigated how tertiary students with a pre-expressed interest in becoming future entry-level intelligence analysts go about searching and gathering information using the WWW.

Background of the Study

Intelligence, traditionally viewed as the art and science (Stephen, 2012) of accessing and evaluating secret information, has been produced since ancient times with the purpose of informing decision makers and leaders across various settings (Richards, 2010). In the early years, covert spies and agents represented the ways of collecting

information that provided rulers and commanders with foreknowledge of terrain, enemy location, strengths and deployment tactics (Olcott, 2012). During the Middle Ages, tradesmen, merchants and European banking houses also started using intelligence as a means of being informed of political and financial practices, as well as local customs of countries and regions of interest, helping them make strategic and tactical decisions (Clauser, 2008; Richards, 2010).

With the end of the fifteenth century came rapid changes and advancements in technology and warfare techniques, marking the beginning of the modern era of intelligence and the development of intelligence doctrines (Clauser, 2008). It was also during this time that Queen Elizabeth started recruiting scholars from renowned universities as researchers and intelligence analysts (Clauser, 2008), paving the path for the future intelligence profession. Most intelligence agencies today primarily recruit tertiary graduates as entry level analysts (Australia Security Intelligence Organisation [ASIO], 2013; Department of Defence Intelligence and Security [DIO], 2013; Federal Bureau of Investigation [FBI], n.d.).

Until recently, information was mainly collected in secret and from clandestine sources, such as satellite images, spies and electronic eavesdropping devices (Prunckun, 2010). Globalisation, mass printing, the introduction of the Internet and the WWW has changed the way in which analysts as well as researchers go about foraging information (Bar-Ilan, 2007; Gill & Phythian, 2012; Olcott, 2012). In the contemporary world, people relatively easily use open sources such as the WWW to access large amounts of information on a daily basis, including information that previously required special security clearance to be accessed as some of this now exists openly on the WWW and in abundance (Bradbury, 2011).

For example, anyone today with an Internet connection can access clear and detailed geographical data through commercial imagery systems such as Google Earth (Mercado, 2004; Steele & Press, 2006). Less than 30 years ago such images were difficult to obtain and only available to analysts with special security clearance. Similarly, social media sites, forums and blogs have made it much easier to map and profile people as private images and information are shared openly on the Web, providing details such as home address, work place, phone number, interests, circle of

friends as well as how they spend their free time. Similar information was previously very difficult to acquire and could take months to map out (Bradbury, 2011).

As such, the art of intelligence is no longer that of accessing scarce information, but rather identifying what is relevant from that what is not (Betts, 2008; Bodnar, 2003; Copeland, 2007; Moore, 2011; Olcott, 2012; Quarmby & Young, 2010). A rich body of knowledge has been developed on intelligence failure, analysis, the role of the analyst, how to improve analysis and the production of intelligence (Canton, 2008; Clark, 2010; Clauser, 2008; Cooper, 2005; Gazit, 1980; George, 2004; Grabo, 2004; Heuer, 2005; Heuer & Pherson, 2011; Lefebvre, 2004; Prunckun, 2010; Richards, 2010; Rodgers, 2006; Walsh, 2011). Little information, however, exists on intelligence open source information foraging, in particular using the WWW.

The intelligence community tends to recruit graduates from various backgrounds as entry-level analysts (ASIO, 2013; DIO, 2013; FBI, n.d.). Understanding how tertiary students gather and use information therefore provides an introductory understanding of how entry-level analysts are likely to gather and use information in the future. Thus, to narrow the gap in knowledge, the current study looked at open source information collection strategies employed by a tertiary student body to solve a specific problem. Participants were recruited from the Bachelor of Counter Terrorism Security and Intelligence course, which has been running for a number of years as a part of the Security Science discipline of a university of metropolitan Perth, Western Australia.

The course information online states, “The degree brings together key aspects of national and international security with an emphasis on terrorism, security and intelligence gathering, analysis and utilisation” (Edith Cowan University [ECU], 2012). Students therefore enter into the course with the objective of going into one of these discipline areas, and some of those who complete this degree will likely work in the intelligence community. Hence, the population from which the sample was retrieved was relevant both academically and vocationally. To this effect, the current study engaged three progressing cohorts (first, second and third year students) in a task asking them to gather information through open sources. This allowed for comparisons across cohorts and the observation of differences in the strategies that were used for the task.

Significance of the Study

As previously mentioned, since the introduction of the WWW, a large amount of information has become openly available and easily accessible to anyone with an Internet connection. Where in the past, the key was finding information that was scares; in the contemporary world, intelligence collection is more focused on sorting relevant information from the available abundance. As intelligence is used to inform decision makers across various domains, basing intelligence on wrong or misleading information may potentially hinder or mislead law-enforcement, financial institutions, governments and military, and can ultimately cost peoples livelihood or lives. Although the Web is used by many people to search for information on a daily basis, relatively little research exists on how this source should be used and what to consider when using it in relation to the intelligence profession.

To this end, the current study investigated how open source information collection skills differ between Counter Terrorism, Security and Intelligence students who are at different stages of a three-year tertiary degree. The study provides insight into the open source information collection skills of potential future analysts as well as indications of whether the degree in which the participants were enrolled in enhances their information collection skills. Hence, the results of the current study are of benefit to both the academic and the intelligence domain.

Purpose and Objectives of the Study

Because of the changing landscape of information and the abundance of open source information existing today, the purpose of the current study was to gain a better understanding of how information is searched for and gathered on the WWW by potential future intelligence analysts. As such, the objectives of the current study were to:

- Identify how students across three different progression stages of a tertiary Counter Terrorism, Security and Intelligence course go about gathering information using the WWW to address a specific intelligence problem.

- Identify significant and/or meaningful differences in the approach to information gathering via the WWW between students of first, second and third year of a tertiary Counter Terrorism, Security and Intelligence course.

Research Questions

As the purpose of the current study was to gain a better understanding of how information is searched for and gathered on the WWW by potential future intelligence analysts, to address objectives and the gap in the existing body of knowledge, the study looked to answer the following research questions.

Principal Research Question

The principal research question was:

- What strategies are employed by students to gather information via open sources at different stages of a tertiary Counter Terrorism, Security and Intelligence course?

Secondary Research Questions

In order to address the principal research question, the following ancillary question was considered:

- Is there evidence of progression in information gathering strategies across the three stages of the tertiary Counter Terrorism, Security and Intelligence course?

Whilst certain differences in information gathering could be assessed quantitatively, other aspects, such as considerations and choices could not. Therefore, a mixed method incorporating quantitative and qualitative approaches was employed.

Structure of Thesis

To achieve the purpose and objectives of the study, and answer the research questions, a five-stage research plan was designed (Figure 1).

Figure 1. Representation of the research process

Chapter One – Introduction to the study sets the scene of the study by presenting the background, stating the significance, purpose, objectives and research questions of the study as well as the structure of the thesis.

Chapter Two – Review of Literature guides the design of the study by outlining core concepts based around the research, expanding on what information exists and identifying gaps in the current body of knowledge.

Chapter Three – Method outlines the research design, sample selection, materials and apparatus as well as details the procedures for data collection.

Chapter Four – Results and Analysis outlines the mixed method data collection results and provides short interpretations.

Chapter Five – Conclusion provides a summary of the study, a discussion of findings, outlines strengths and weaknesses as well as directions for future research.

Chapter 2: Review of Literature

Introduction

This section presents the reviewed literature informing the study. The review firstly considers the main themes that emerge from the literature with regards to how intelligence analysts gather and use Open Source Intelligence (OSINT). This part looked to define intelligence and the intelligence process, whilst focusing on identifying key trends and issues with information collection. The aim of this section is to identify how intelligence analysts utilise open sources such as the WWW to search for information in a timely and efficient manner to meet information deadlines. To address the existing gap in knowledge regarding the collection of OSINT using the WWW, and as the intelligence profession primarily recruits recent university graduates as entry-level analysts, the review secondly looked at the main themes that emerge from the literature with regards to how tertiary students forage open source information using the WWW.

Thus, the second part of the literature review focused on defining the Web and identifying key issues as well as considerations with using the WWW as a source to forage information. Essentially, by establishing how students use the Web to search for and gather information, deeper understanding of how entry-level analysts are likely to go about foraging information on the Web will be gained. Hence, the aim of this section was to extract and define what Web information literacy is and what strategies are employed by tertiary students in an efficient approach to searching for information on the Web. The reviewed literature informed the methodology for the study towards elucidating common trends in efficient and effective information researching for intelligence analysts using the WWW.

Intelligence, Analysis and the Analyst

Intelligence has been produced since ancient times and forms a fundamental part of informing and aiding decision making on strategic, operational and tactical levels (Olcott, 2012). To understand how intelligence analysts gather and use electronic open source information, it is first of importance to have a basic understanding of what intelligence is, how it is produced and who uses it. Traditionally intelligence was utilised by the government, national security, foreign policy, law enforcement and defence domains (Richards, 2010). In the contemporary world, however, intelligence has extended far beyond these areas and also become a critical function of commercial business domains such as the Australian tax office, state fisheries departments and natural resource organisations (mining, oil and gas industries).

Because intelligence is utilised in such a large variety of domains, as mentioned by Walsh, "Defining 'intelligence' in a generic sense . . . is contentious, different perspectives are included or excluded depending on the view of the scholar or practitioner" (2011, p. 9). As such, there are many different definitions of intelligence. Scott and Jackson suggest, "In most contemporary analyses, intelligence is understood as the process of gathering, analysing and making use of information" (2004, p. 141). However, most professionals agree that intelligence is both a product and a process (Gill & Phythian, 2012; Quarmby & Young, 2010).

The intelligence process is predominantly referred to as the intelligence cycle, a sequence of interrelated processes that represent the different stages through which the intelligence product is produced (Cooper, 2005; Richards, 2010; Walsh, 2011). Essentially, these stages comprise of planning and direction, collection, processing, analysis and dissemination. Briefly explained, the direction phase interprets what the decision maker needs, directs and sets the information requirements for the collection phase, which aims to search and gather data from various sources to fill the existing information gaps (Gill & Phythian, 2012).

What sources are used to gather information depends on the nature of the problem, budget, deadline for dissemination and the existing "unknowns". However, in general the literature promotes using all-source gathering (Bean, 2011; Gibson, 2004;

Mercado, 2004; Olcott, 2012; Steele & Press, 2006), where the main sources used are Human Intelligence (HUMINT), Signals Intelligence (SIGINT), Imagery Intelligence (IMINT), Measurement and Signatures Intelligence (MASINT) and OSINT (Gill & Phythian, 2012). All new gathered data are processed and evaluated together with its source in order to establish reliability and validity in preparation for analysis. Once data have been gathered, processed, evaluated and collated with existing data, an intelligence analyst develops the end product by synthesising and making sense of the information, and disseminates it to the decision maker in a timely and appropriate manner (Gill & Phythian, 2012; Richards, 2010).

Although these phases may appear simple and separate, in reality they run simultaneously, interchangeably and in different scales depending on information requirements and the complexity of problem (Gill & Phythian, 2012). Context plays an important part in intelligence production as everything is intertwined. Essentially, the product is only as good as the analyst's capability to make sense of uncertainty, apply knowledge and critically analyse gathered information in relation and relevant to a specific problem (Canton, 2008; Gill & Phythian, 2012; Lefebvre, 2004; Moore, 2011).

Some scholars argue that, although analysis is their primary role, the analyst also directs and participates in intelligence collection as well as disseminates the end product to the decision maker (Canton, 2008; Cooper, 2005; Gill & Phythian, 2012; Lefebvre, 2004; Quarmby & Young, 2010; Richards, 2010). As such, many scholars see intelligence analysis and the role of the analyst as the most important part to the production process of intelligence (Canton, 2008; Cooper, 2005; Gill & Phythian, 2012; Lefebvre, 2004; D. T. Moore, Krizan, & Moore, 2005; Richards, 2010; Walsh, 2011). As expressed by Lefebvre, "Why would intelligence agencies collect mountains of data if not to make sense of it and provide policymakers with their best judgment as to its meaning and implication? There is simply no point in collecting data to sit idle" (2004, p. 235).

Even so, intelligence analysis is based on gathered information and the accuracy of predictions depend on the completeness, relevance, validity and accuracy of the data analysed just as much as the capabilities of the analyst (Lefebvre, 2004; Rodgers, 2006). It is therefore of extreme importance that sufficient and relevant information is gathered

as incomplete and irrelevant data or, a more potent contemporary issue, too much data gathered resulting in information overload may lead to inaccurate predictions and, ultimately, intelligence failure (AFCEA Intellgence Committee, 2005). As such, it may be argued that information collection failures are the most damaging to the production of intelligence and therefore, how information is gathered should be given more attention.

With globalisation, mass media and the introduction of the WWW came drastic changes in the open source information landscape (Betts, 2008; Gill & Phythian, 2012; Lefebvre, 2004; Moore, 2011; Quarmby & Young, 2010), making large amounts of information openly available to the general public. In turn, this led to a shift in the focus of information collection, where achieving a high volume of information was previously important because volume was seen to likely provide that critical piece of the jigsaw that answers the question (Betts, 2008). However, information today is so easily accessible that volume is no longer seen as a key measure, volume is now easy to get. The issue now has become identifying that which is relevant within the noise and overflow of information (Betts, 2008; Bodnar, 2003; Copeland, 2007; Lefebvre, 2004; Moore, 2011; Olcott, 2012; Quarmby & Young, 2010; Richards, 2010).

As such, there is a great need to better understand how intelligence analysts go about searching for and gathering information using open sources such as the WWW. Although other sources such as HUMINT, IMINT, MASINT and SIGINT exist and may be considered equally important (Steele, 2007), it is beyond the scope of the current study to discuss them in any further detail.

Open Source Intelligence

According to Steele, "Open source intelligence, or OSINT, is unclassified information that has been deliberately discovered, discriminated, distilled and disseminated to a select audience in order to address a specific question" (2007, p. 129). What is meant by "unclassified information" is information that openly exists in a legal manner (Gibson, 2004), such as official government and business documents, speeches and radio/TV broadcast, as well as information published on the Internet, in newspapers and journals (Gill & Phythian, 2012). Although OSINT has been used by both military and government intelligence organisations since World War II (Glassman & Kang,

2012), the last 25 years has brought changes that have greatly influenced the open source information domain (Bean, 2011).

With the reform of information communications technologies, such as the Internet and the WWW, certain data became openly and legally available that previously required security clearance to be accessed (Gill & Phythian, 2012; Hulnick, 2006). For example, to establish the identity or build a profile of a person of interest, the intelligence analyst was previously required to access personal records and classified satellite images of where the person was living and working (Gill & Phythian, 2012). However, today, with as little information as the full name, address of where they live and work, it is possible to map out most likely rout that the person takes to work, where they would shop for groceries, where their kids may go to school as well as what interests they have and how they spend their free time (Bradbury, 2011). All this information is now openly available on the WWW through social media sites, forums, blogs and Google maps, as well as through phone GPS systems (Appel, 2011).

Not only certain classified data became openly available, the WWW fast became the one stop shop for information, hosting newspapers, journals, online books, business and government Websites as well as personal Websites, blogs and social media profiles (Appel, 2011). Although some of these sources were highly relevant and used before, as suggested by Bradbury:

> These days, such sources are still highly relevant, but there is far more of that information to sift through. And the availability of other kinds of information, such as metadata in documents and social networking data, has made open source intelligence even more useful, while also making it harder to manage. Suddenly, sourcing publicly available information has become like drinking from a firehose. But it is also a key tool for everyone from law enforcement through to merger and acquisition teams, headhunters, and anti-fraud departments in private organisations. (2011, p. 5)

A greater urgency was placed on OSINT after the introduction of the WWW as it made a seemingly unrestricted and constantly growing amount of information openly available to everyone with an Internet connection (Glassman & Kang, 2012). Thus, changing the information landscape and shifting the intelligence profession from that of being experts at finding scarce information, to that of finding and identifying relevant pieces of information in an endless and rich information environment (Steele, 2007). False and deceptive information exists on the Web, which can make it very difficult to verify and find relevant information within the large amount of noise. However, in addition to its broad coverage, what makes OSINT and Web information imperative to the intelligence profession is, as the information is unclassified, that the intelligence produced can easily be shared between entities without compromising sources (Bean, 2011).

Even so, academics and intelligence professionals have developed a substantive body of knowledge on intelligence, intelligence analysis, failures, the changing intelligence environment and improvement possibilities over the last 15 years (Canton, 2008; Clark, 2010; Clauser, 2008; Cooper, 2005; Gazit, 1980; George, 2004; Gersh, Lewis, Montemayor, Piatko, & Turner, 2006; Gill & Phythian, 2012; Grabo, 2004; Heuer, 2005; Heuer & Pherson, 2011; Lefebvre, 2004; Moore, 2011; Prunckun, 2010; Richards, 2010; Rodgers, 2006; Walsh, 2011). There are also number of studies that look at implementing computer based models to search for, gather and/or sort open source information (Bodnar, 2003; Camacho, Aler, Borrajo, & Molina, 2006; Gersh et al., 2006; Hulnick, 2006; Moore, 2011). Although interesting, this topic is not of relevance to the current study and will therefore not be discussed further.

Previously mentioned, research regarding open source intelligence collection using the WWW is relatively scarce and thus, difficult to find. Although a number of information pieces are published which discuss OSINT (Bean, 2011; Bradbury, 2011; Olcott, 2012; Steele, 2007; Steele & Press, 2006), these primarily focus on the "what" and not the "how", discussing the changing information landscape, what it means, the importance of OSINT and how this source should be implemented through all intelligence domains. Thus, providing very little detail on how intelligence analysts actually go about foraging information on the WWW. Even so, a number of scholars

provided some knowledge on the collection process, making it possible to define how analysts in general go about gathering open source information.

From the contemporary literature, three main considerations to open source information collection have been identified. Firstly, the analyst needs to understand the decision makers problem and what information is needed to address it, as such develop a picture of what is required (Canton, 2008; Gill & Phythian, 2012; Moore, 2011; Prunckun, 2010). Secondly, with this understanding the analyst is then able to identify where relevant information may come from, how best to go about gathering it, and create a collection plan that directs search terms and places (Clauser, 2008; Moore, 2011; Prunckun, 2010; Quarmby & Young, 2010). Lastly, when searching for information, the analyst must be able to relatively quickly sort information that is relevant, filter out the noise and focus on quality of information rather than quantity (Betts, 2008; Clauser, 2008; Copeland, 2007; Prunckun, 2010).

Although the above description provides a general overview of how intelligence analysts go about using the WWW to forage information, there is a clear gap in the knowledge regarding the details of the "how" of Web gathered OSINT. As such, there is a need to further understand the WWW, what Web information literacy means and how information can be efficiently searched for and gathered using the WWW.

The World Wide Web

According to Bar-Ilan, "The Web has become a major source of electronically stored information in the developed world, answering many of people's information needs in their everyday, personal, and professional lives." (2007, p. 910). Trends over the last ten years suggest that the amount of information being shared on the Internet is growing at an exponential rate, where approximately 800 million pieces of information were available in the year 2000 (Agosto, 2002). In just five years this figure grew to an estimated 11.5 billion (Gulli & Signorini, 2005). As such, there is a growing need to better understand what the WWW is, how it is used to search and gather electronically stored open source information, and the issues and limitations that arise from its capacity and storage architecture.

Downes (2007) describes the WWW as the largest information store in the world, allowing pages of information to be shared between people by an interconnecting series of computers via the Internet architectures. From the Web, information in the form of documents, websites, videos, images, articles, and news magazines, amongst other formats, can be retrieved from any other Internet connected computer (Downes, 2007). Although media groups, organisations, governments, commercial businesses and academic establishments upload and share enormous volumes of information on the Web, Downes (2007) highlights that a large amount of information is also shared by individuals who create personal websites or blogs expressing their own thoughts and knowledge on any given topic.

Globally any person at any time can upload information to the Web in the form of personal Websites, blogs and forums without it being validated, evaluated, peer reviewed, or going through any other prior vetting of the source supporting the information before being released to the general public (Burkhardt, MacDonald, & Rathemacher, 2010). This is in stark contrast to the information security principles of Confidentiality, Integrity and Availability (CIA) (Stamp, 2011). Doyle and Hammond (2006) emphasise that this is a significant limitation of Web based information, as most people searching for and gathering information require the truth pertaining to a topic of interest, thus seek valid information in which they feel is authentic and therefore justified in believing.

Accordant with the CIA model, identifying the reliability of an information source or piece was perhaps less difficult in the past as researchers saliently utilised and searched through printed documents, journals and books, often stored in libraries (Neely, 2006). Validity was not of concern as most information had been peer reviewed and/or edited by at least one, usually two recognised experts prior to publishing (Doyle & Hammond, 2006). However, in contrast to such controls, textual changes on the Web are difficult to track (Choo, Turnbull, & Detlor, 2000; Wolfe, 2001) as these are not required to be marked and no edition number or edited stamp is required. Therefore, researchers are often not able to identify whether information has been changed and if the information is authentic, thus compromising integrity. In addition to document integrity being difficult to track, de Kunder (2012) points out that the sheer volume of

information available on a daily basis is both overwhelming and constantly changing, thus compromising Availability.

In the last year alone, the number of indexable pages openly available on the WWW has ranged between 10 to 20 billion pieces per day (de Kunder, 2012). This highlights that, on top of new information being constantly uploaded to the Web, daily previously available information is also disappearing (Lewandowski, 2008). Although Lewandowski's figures are now out-dated, they provide some indication as to how vastly the WWW information landscape is changing:

> Estimating the results . . . there are about 320 million new pages every week. About 20 percent of the web pages of today will disappear within a year. About 50 percent of all contents will be changed within the same period. The link structure will change even faster: about 80 percent of all links will have changed or will be new within a year. (Lewandowski, 2008, p. 818)

Thus, there are a number of concerns related to information on the WWW. Even though the Web is the largest information store in the world, its sheer volume can overwhelm even the most experienced of researchers. In addition, some of the information existing on the Web stem from personal opinions as anyone can upload/change/delete information at any time without prior reputable scrutiny, or notifications of change. Therefore it can sometimes be difficult for researchers to identify reliability of information and sources, as well as track changes made to information. This highlights an area of significant importance in understanding the volume, relevance and time required to sift through retrieved results when searching for information on the WWW.

How to conduct searches on the Web

When people search for information on the Web, results can be retrieved in one of three ways, by typing in its known WWW address, by following a hyperlink, or by using a search engine (Alexander & Tate, 1999; Downes, 2007). Search engines are most widely used by undergraduate students as a means of searching for information sources on WWW (Biddix, Chung, & Park, 2011; Jones, Johnson-Yale, Millermaier, & Pérez, 2008; Judd & Kennedy, 2010; Metzger, Flanagin, & Zwarun, 2003). This is

arguably due to their aiding with finding and retrieving information without having to know, or enter, a full Web address, or having access to a specific hyperlink (Herring, 2011). Thus, only search engine information searches will be further discussed as part of the study literature review.

According to a number of researchers, some of the most widely used search engines are: Google, Yahoo!, Windows Live (MSN), Ask and Bing (Bar-Ilan, 2007; de Kunder, 2012; Doyle & Hammond, 2006; Gulli & Signorini, 2005; Lewandowski, 2008; Spink, Jansen, Kathuria, & Koshman, 2006; Uyar, 2009). By entering words or sentences into the search engines as search terms, information is retrieved and displayed to the researcher in rank order. How this rank order is determined depends on a number of criteria, including search terms used and mechanics of the specific search engine (Lanning, 2012; Ware, 2001). As explained by Eliopoulos and Gotlieb, "Essentially, the effectiveness of a search engine defines the scope of what the user is "allowed" to find. If a search engine is set up poorly, the users may never find what they are looking for" (2003, p. 42). Some search engines require sources to pay money for higher ranking, others use algorithms that take a number of criteria into account when ranking, such as number of sources linking to the page, freshness of page, and/or frequency of visits to page (Doyle & Hammond, 2006; Eliopoulos & Gotlieb, 2003; Lewandowski, 2008; Spink et al., 2006).

However, even when using an appropriate search engine, entering too narrow, too broad or irrelevant query language as search terms will generate irrelevant information results (Lanning, 2012; Ware, 2001). This can make it very difficult and time consuming for the researcher to find their relevant information. More often than not, even with appropriate search terms, far too many results are produced and displayed to the end user, making it almost impossible to sift through all information pieces (Nachmias & Gilad, 2002). As such, important or relevant information may often be missed as many researchers only go through the first few numbers of pages listing the search results (Gulli & Signorini, 2005).

Even though search engines simplify the search process by going through vast amounts of information to find and retrieve relevant results on the behalf of the researcher, there are a number of concerns and/or limitations highlighted by the

literature that should be taken into account when using them. Such include the way the search engine ranks the returned results, their sensitivity to how the researcher structures their search query and the overwhelming results that may be returned even with appropriately structured searches. Thus, reinforcing the area of significant importance to understand the volume, relevance and time required to sift through search results, and adds the importance to understand how to appropriately use Web search tools in order to efficiently retrieve relevant information.

Previous findings and collection barriers

Considerable effort has been devoted towards researching how students go about searching, gathering and evaluating information on the WWW (Braasch et al., 2009; Gross & Latham, 2007; Judd & Kennedy, 2010; Kim & Sin, 2011; Metzger et al., 2003; Thompson, 2003), and to develop various Web information literacy teaching methods for librarians and teachers (Alexander & Tate, 1999; Burkhardt et al., 2010; Herring, 2011; Neely, 2006; Ware, 2001). However, the salient focus of such research has been directed towards school children and young adults (Agosto, 2002; Bilal, 2000, 2001, 2002; Bilal & Bachir, 2007; Braasch et al., 2009; Britt & Aglinskas, 2002; Coiro, 2011; Dresang, 2005; Fidel et al., 1999; Hoffman, Wu, Krajcik, & Soloway, 2003; Large & Beheshti, 2000; Tu, Shih, & Tsai, 2008). Minimal studies have focused towards adults or tertiary level participants searching the WWW for information (Britt & Aglinskas, 2002; Doyle & Hammond, 2006; Judd & Kennedy, 2010; Maybee, 2007; Nachmias & Gilad, 2002; Wiley et al., 2009).

The existing literature at large focused on evaluating how they approached the Web when searching for information in relation to a specific question or assignment, either focusing on search terms and search engines used (unless specified) (Bilal, 2000, 2002; Fidel et al., 1999; Hoffman et al., 2003; Large, Beheshti, & Breuleux, 1998; Thompson, 2003; Tu et al., 2008), and/or time spent on the given task by each participant (Bilal, 2001). Such studies also considered the amount of information viewed or scanned by each participant (Judd & Kennedy, 2010), and/or how participants evaluate reliability (Braasch et al., 2009; Britt & Aglinskas, 2002; Coiro, 2011; Kim & Sin, 2011; Large et al., 1998; Metzger et al., 2003) of online information. Further, some studies have focused on what participants thought of using the Web as a

research tool to gather information (Agosto, 2002; Fidel et al., 1999; Large & Beheshti, 2000). Although different approaches and methods were utilised, similar findings were identified by most researchers to that of students' collection barriers on the Web.

One of the main findings of the current literature was that the sheer volume of information existing on the Web often overwhelmed and distracted individuals when searching for specific information (Bowler, Large, & Rejskind, 2001; Doyle & Hammond, 2006). As a result of the large volume of information often returned by search engines, most students were found to skim through information too rapidly to find anything of relevance, leading them to often miss vital information (Bowler et al., 2001). In addition, studies showed many retrieved a large volume of irrelevant information when using search engines, and therefore less experienced Web users were often required to go through a vast amount of information to find a few relevant pieces of information (Fidel et al., 1999; Large & Beheshti, 2000). Yet in contrast to this, more experienced Web searchers could identify relevant information more efficiently (Britt & Aglinskas, 2002; Coiro, 2011; Nachmias & Gilad, 2002). This suggests that experience, and as such contextual education helps in the retrieval of relevant information on the Web.

Findings also indicated that it was difficult for most students to identify appropriate search strategies, either using too specific or too broad search terms, which influenced the amount of irrelevant information being generated (Bilal, 2001; Bowler et al., 2001; Braasch et al., 2009; Large et al., 1998; Nachmias & Gilad, 2002; Tu et al., 2008; Wiley et al., 2009). A salient concern that emerged from these studies was the difficulty faced by students when trying to evaluate the validity of information from the WWW (Bowler et al., 2001; Braasch et al., 2009; Bråten, Strømsø, & Britt, 2009; Brem, Russell, & Weems, 2001; Britt & Aglinskas, 2002; Doyle & Hammond, 2006; Hirsh, 1999; Hoffman et al., 2003; Lorenzen, 2002; Tu et al., 2008). As a result, nearly all findings expressed the need for educating young students in how to use, search for and gather information on the Web (Agosto, 2002; Bilal, 2000, 2001, 2002; Bowler et al., 2001; Braasch et al., 2009; Coiro, 2003, 2005, 2009, 2011; Coiro & Dobler, 2007; Fidel et al., 1999; Hirsh, 1999; Large & Beheshti, 2000; Nachmias & Gilad, 2002; Wiley et al., 2009).

Efficient Web information searching

Hoffman, Wu, Krajcik and Soloway (2003, p. 325) describe Web information seeking as "a special case of problem solving . . . in which learners recognize and interpret an information problem, establishing a plan of research, conduct the search, evaluate the result, and use information to solve a problem". This is referred to by Maybee (2007) as "the process" of information seeking. There are many ways to search for, forage and gather electronically stored information. However, research suggests that to search for, forage and gather information efficiently and successfully on the WWW, specific knowledge and skills are required (Alexander & Tate, 1999; Burkhardt et al., 2010; Doyle & Hammond, 2006; Fraenkel & Wallen, 2006; Lanning, 2012; Neely, 2006; Ware, 2001; Wolfe, 2001). This is consistent with the view of Herring, that "people who are effective web searchers have developed a regular approach to searching, which can be seen as a set of rules of habits" (2011, p. 27).

Research suggests that for a student to be efficient and successful in searching for information on the Web, there are four main strategies that they must adopt as part of their approach. First, they are required to clearly identify and become familiar with their research topic in order to recognise and define information requirements, identify common terminology and key concepts relevant to the research topic, and to identify type of information needed to respond to information requirements (Doyle & Hammond, 2006; Herring, 2011; Lanning, 2012; Neely, 2006). This strategy provides the researcher with a clear understanding of what they are searching for, type of information they require and where best to find it.

Second, the researcher is required to identify and select specific search engines that will allow them to access required information (Neely, 2006), and to understand their underlying search principles in order to break down key concepts into searchable key words/phrases and effectively construct these into search strategies using relevant language structures (Boolean, controlled vocabulary, natural language etc.) for the selected search engines (Burkhardt et al., 2010; Lanning, 2012; Ware, 2001). This strategy provides the researcher with the skills necessary to efficiently construct and conduct searches, and to filter and refine search strategies to yield more relevant search results.

Third, the researcher is required to evaluate authority, accuracy, currency, objectivity and coverage of information and sources (Alexander & Tate, 1999; Johnson & Christensen, 2012; Ware, 2001). This strategy provides the researcher with the ability to effectively select and retrieve only such information that is usable and relevant to information requirements. Fourth, and lastly, the researcher is required to organise and synthesise gathered information in order to effectively respond to and, if necessary, re-evaluate the information needs (Neely, 2006). This strategy aids the researcher with communicating their findings efficiently and in a meaningful way.

As the current study aims to identify how future intelligence analysts go about gathering information using open sources such as the WWW, focus is placed on looking at the first two strategies of efficient Web information searching. Although relevance and usability of gathered information will be considered, because of time limitation the study will not place focus on how students evaluated sources and information, or why they selected the specific information that they gathered. Whilst the fourth strategy is highly relevant to Web information literacy, as the current study looks at information collection only, students will not be required to organise, synthesise or communicate gathered information. Thus, it is beyond the scope of the current study to look at this strategy any further.

Summary

The ability to understand the universe of information (the breadth of information), develop a strategy for how to best gather relevant information and narrowing the domain, knowing where to look, how to look as well as recognise the significance of what is found when searching for information is what intelligence collection is about. Although these considerations are not specific to just OSINT, interestingly they are almost identical to the first three strategies and skills identified as necessary when foraging information using the WWW. Both search methods highlight the need to establish information context and identify information requirements, develop appropriate search strategies for where best to find the information and how to find it, and to be able to identify and evaluate relevant information.

Whilst the literature on Web foraging highlights a number of concerns with searching for information on the WWW, which includes that reliability of information is more difficult to identify. This is exacerbated by the lack of bibliographic control on the Web and that the volume of information is rapidly growing as well as constantly changing. Furthermore, information on the WWW can be overwhelming and is constantly changing, impending the reliability of retrieval, requiring people to use search engines to find and gather information. As such, information can be missed and more time needs to be spent on developing search strategies and identifying and evaluating the sources behind the upload of information. These concerns are also similar to what has been identified by the intelligence community.

The literature also details what makes an efficient and successful Web searcher, which includes the ability to clearly establish information context, use the right search tools and search strategies to retrieve relevant information to given context, critically evaluate the usability of information and sources against set criteria, and to organise and synthesise gathered information to address and respond to information needs. Furthermore, explaining that educating students and aiding them with the development of such skills will enhance Web information literacy of students. The literature found on intelligence collection, however, provides very little detail on such information in relation to OSINT. As such, there is a gap in knowledge on how analysts actually go about gathering open source information using the WWW and how the skills identified as necessary are developed.

Therefore, the purpose of the current study was to undertake an experiment with three consecutive year levels of tertiary student from a Counter Terrorism, Security and Intelligence course of a university in metropolitan Pert, Western Australia. The study employed a mixed method approach, qualitatively measuring number of search terms used, amount of information viewed and amount of information gathered by each cohort. To provide deeper meaning and assess the congruence between participants´ subjective experiences and empirically observable strategies, the students were also asked to quantitatively list what strategies they employed when addressing the task. Thus, addressing the principal research question: What strategies are employed by

students to gather information via open sources at different stages of a tertiary Counter Terrorism, Security and Intelligence course?

The data and findings were further compared between cohorts to measure skill improvements, thus addressing the second research question: Is there evidence of progression in information gathering strategies across the three stages of the tertiary Counter Terrorism, Security and Intelligence course? In addition, an expert in intelligence analysis and information collection was employed to identify whether any progression exists between cohorts in relation to relevance and usability of gathered information.

Chapter 3: Method

Introduction

This chapter presents the research methodology and design utilised to address the existing gap in literature and answer the principal and secondary research questions. The research design is discussed, followed by a small pilot, participants, and materials and apparatus. Once these have been presented, the procedures of the current study are be outlined, followed by a short conclusion of the chapter.

Research Design

As the purpose of the current study was to gain a better understanding of how potential intelligence analysts search for and gather information on the WWW, the current study employed an experimental mixed method design. This approach allowed data to be collected quantitatively and qualitatively, providing richer information on the strategies employed by participants (Creswell, 2009; Johnson & Christensen, 2012). This design was deemed an appropriate data collection method for the experimental nature of the study and for addressing the research questions (Plano Clark & Creswell, 2008). By looking at how students with a pre expressed interest in becoming intelligence analysts go about searching and gathering information on the WWW, the research provided important insight into what potential Web searching skills future analysts may have and whether a tertiary degree influences the development of these skills.

A task was designed where participants were asked to gather information relevant to a real time problem (the independent variable). This provided the opportunity to measure three dependent variables quantitatively; amount of information gathered (operationalised as the number of pieces of information gathered), number of search terms used (operationalised as the number of search terms utilised to search for information) and number of clicks used by each participant (operationalised as the number of pieces of information accessed), in three cohorts of tertiary students. Where the first year cohort represent students who are enrolled in their first intelligence unit

and, as such, have little to no experience of intelligence information gathering. Second year participants represent students with some prior knowledge as they have completed the first year of their three-year tertiary course. Participants completing their last year of the course were classified as third years.

After completion of the task, participants were asked to answer a qualitative open-ended question that allowed a deeper understanding of strategies and considerations employed by participant when engaging the task. As such, providing an opportunity to identify information gathering skill differences between cohorts as a function of a three-year tertiary course. Participants were also asked to fill out a basic demographic sheet, so that the data could be analysed to assess demographic effects. This sheet included questions regarding the person's individual characteristics, such as age, gender and relevant work experience.

Pilot Study

Prior to the actual study, a small pilot was conducted with three participants to assess face validity of the instructions and the workings of software that was to be used (Plano Clark & Creswell, 2008). The pilot allowed for identification of some necessary minor refinements of the task instructions, which essentially resulted in participants being provided with a more detailed description of what was expected of them. As a result of the pilot, a sheet containing examples of what participants' were expected to copy and paste into the Word collection file were developed and included in the instructions for clarification purposes (Appendices D). In addition, it was identified that the suggested screen capture software (CamStudio) that was to be used to record each participant's movement on the Web did not function reliably, rendering it redundant. Therefore a print screen of each participants Web browser history was to be generated and saved onto a separate Word file to replace the screen capture software and limit potential data loss.

Participants

The sample consisted 40 students, ranging in age from 17 – 52 years (M = 24.88, SD = 8.13), currently enrolled in a three-year Bachelor of Counter Terrorism, Security and Intelligence course. Although it is recommended to use random sampling of participants (Cohen, Manion, & Morrison, 2007; Creswell, 2009), as there were limited numbers all students enrolled in the degree were invited to participate, exhausting the sample population within some of the cohorts. The sample comprised 1 female and 10 male first year students (n = 11, 27.5%), 1 female and 15 male second year students (n = 16, 40%), and 2 female and 11 male third year students (n = 13, 32.5%). The average age of participants was 27.09 years for the first year cohort (SD = 10.98), 24.88 years for the second year cohort (SD = 9.34) and 23.00 years for the third year cohort (SD = 4.38).

Materials and Apparatus

The current study comprised an information letter detailing the aim of the study (see Appendices A), a consent form that each participant was required to sign prior to commencing (see Appendices B) and instructions detailing the real time problem and task (see Appendices C). Example images clarifying what information should be copied and where to paste the information into the Word document were also provided together with the instructions (see Appendices D). A demographic sheet containing six questions concerning participants' age, gender, English as first language, educational level, prior relevant experience and year of course that student was enrolled in was also provided (see Appendices E) together with a qualitative question (see Appendices F) asking students to detail in dot format strategies employed when performing the task.

There were two types of materials used to collect data in the current study, information gathered and search terms used by each participant, together with a history print screen were recorded onto Microsoft Word documents. The strategies employed when approaching the problem and the demographic details of each participant were recorded onto separate sheets of paper.

The actual study was conducted in two computer laboratories of a university in metropolitan Perth, Western Australia. Both computer labs were similar in nature concerning ergonomics, lighting, air conditioning and outlay. However, they had slightly different computers and varied in sizes. Please see Table 1 for more details. Whilst, in reality, intelligence analysts perform searches on demand from different settings and at all possible variances of time and day, these include similar settings as within the computer labs used for the current study. Hence, external validity was addressed. A stopwatch was also used during the study.

Table 1

Information about labs and computers used in the study

Characteristic of Computers	Computer Lab 1 (n = 33)	Computer Lab 2 (n = 6)
Make and Model	Custom Built	Dell OptiPlex 9010
Size of monitors	Samsung 24 inch	Samsung 22 inch
Operating system	Windows 7 Pro 64 bit	Windows 7 Enterprise, SP1
Ram	16GB	4GB
Processor	ADM64 8 core, 3.2 GHz Processor	Intel core i5, 3.2 GHz Processor
Internet Explorer	Version 8	Version 8
Internet connection	Auto negotiated up to 1gb	Auto negotiated up to 1gb
NO. of computers	20	25

Note. SP1 = service pack 1; GB = gigabytes; GHz = gigahertz

Procedures

Prior to commencing the recruitment process, approval was sought and given by the Ethics Committee of the Faculty of Computing, Health and Science of Edith Cowan University. As one of the supervisors of this research lectured and coordinated intelligence analysis units at the time, the opportunity to ask students to volunteer for the current study during assigned lecture hours was provided. Students interested in participating in the study were given an information sheet (see Appendices A). The supervisor also aided with sending out two recruitment emails to all students enrolled in

the targeted course, providing information about the study, asking for volunteers and specifying study times, in order to reach students not present at the time of lecture. The study times were predominantly conducted during students' existing tutorial times, which provided an opportunity to maximise participants' number.

The actual experiment ran over 50 minutes, which is congruent with usual tutorial time. After 5 minutes of introduction and signing informed consents, each participant was given a task sheet, listing the real time problem and instructing participants to copy search terms, results each search term generated and the Uniform Resource Locators (URLs) together with the part of the information found relevant to on to a Microsoft Word document (see Appendices C). Prior to the commencement of the study, participants were given the opportunity to withdraw and ask question regarding the task. Once all questions were answered, the participants were given 30 minutes, timed by the researcher with a stop watch, to search for and gather information through open sources, using computers connected to the Internet, to solve a real time intelligence problem. Once search time ended, participants were given 10 minutes to fill out the demographic sheet (see Appendices E) and answer one qualitative question (see Appendices F).

This provided each participant with an equal amount of time to engage the task and answer the question. The nature of the real time intelligence problem was similar to that of what students attending this tertiary degree typically are presented with (see Appendices C). As such, the task and problem used did not place participants in any foreseeable risk in terms of personal distress, or otherwise control for potential confounds related to time of task administration.

Initially, the study was set to run over four sessions at a designated computer laboratory over the course of two days. However, in an effort to increase number of participants for the first and third year cohorts, by the end of day two the study was extended and set to run for additional three weeks in a second computer laboratory. As such, providing a total of eight sessions over a four-week period for the students to participate.

Chapter 4: Results and Findings

Introduction

This chapter presents the analysis and results of the data. As the current study utilised a mixed method approach to address the principal and secondary research questions, data were analysed in two ways. The continuous dependent variables of 'number of search terms used', 'amount of information gathered', and 'number of clicks used' by each participant were examined statistically across the tree cohorts. Thematic analysis was used to analyse participants' responses to the question: "in dot form, please list strategies you employed to achieve the given task". The analyses used allowed for observations of differences between cohorts as well as assessment of the congruence between participants subjective experiences and empirically observable strategies. Quantitative results will be reported first, followed by the findings and the interpretations of the qualitative data.

Quantitative Results

No missing data were identified prior to analysis when screening all 40 Word data collection files and demographic sheets. All data were entered into the Statistical Package for the Social Science (SPSS) for Windows (version 22.0) for screening and analysis purposes. Various SPSS procedures were used to examine the three dependent variables within each cohort for missing values, univariate and multivariate outliers, and violations of normality as well as homogeneity of variances. No missing values were detected. Alpha was set at .05.

No multivariate outliers were identified using Mahalanobis distance with $p <$ .001. Four univariate outliers were detected across the year levels. Two outliers were detected for the second year cohort on the dependent variable of 'number of search terms used', one for the first year cohort on the dependent variable of 'amount of information gathered' and one for the third year cohort on the dependent variable of 'amount of information gathered'. In order to sacrifice minimum variance (Tabachnick & Fidell, 2007), the four univariate outlying values were adjusted by changing their

values to the most extreme value allowed in the direction within the 95% confidence interval.

After adjusting the univariate outliers, Shapiro-Wilk test, used as $N < 50$, identified that the assumption of normality was violated for 'number of search terms used' for the first year cohort and 'number of clicks used' for the third year cohort. Levene's test identified that the assumption of homogeneity of variances was violated for the 'number of search terms used'. Pearson Correlation analysis was conducted to assess the correlation between the three dependent variables (see Table 2). Although there was a statistically significant positive correlation between the 'number of search terms used' and 'amount of information gathered', $r = .35$, $N = 40$, $p = .027$, this was not a strong correlation as it indicated only 12% shared variance between the two variables.

Table 2

Correlation Among Study Variables

	Amount of Information Gathered	Number of Search Terms Used	Number of Clicks Used
Amount of Information Gathered	_____	.35*	0.29
Number of Search Terms Used		_____	0.28
Number of Clicks Used			_____

Note. Correlation marked with an asterisk (*) was significant at $p < .05$.

Because there were no strong correlations between variables and there were violations of both normality and homogeneity of variance assumptions, it was not appropriate to run parametric tests such as the Multivariate Analysis of Variance (MANOVA) or Analysis of Variance (ANOVA). As suggested by Tabachnick and Fidell (2007), the more appropriate option to examine the data was to run nonparametric tests that are robust against these violations. As there were three independent cohorts, Howell (2008) suggests running a one-way Kruskal-Wallis test to evaluate differences among the cohorts. This found a statistically significant result, with $\chi^2(2, N = 40) = 6.09$, $p = .048$, for the 'number of search terms used'.

To identify the location of the statistical significance, three nonparametric Mann-Whitney U pairwise post-hoc tests were conducted (Howell, 2008). To reduce the likelihood of familywise error, a Bonferroni adjustment was applied yielding an alpha of .017. These revealed that a statistically significantly higher number of search terms were used by the second year cohort compared to the first year cohort, $z = 2.35$, $p = .017$. No other statistically significant differences were identified. Descriptive statistics are presented in Table 3.

Table 3

Mean Scores for Amount of Information Gathered, Number of Clicks Used, and Number of Search Terms Used

		Amount of Information Gathered		Number of Clicks Used		Number of Search Terms Used	
Participant Year Level	*n*	M	*SD*	M	*SD*	M	*SD*
First Year	11	6.09	3.21	14.45	6.25	4.55	1.69
Second Year	16	8.56	3.74	17.13	6.29	8.19	4.35
Third Year	13	8.54	3.41	18.31	7.73	7.54	4.27

Note. $N = 40$.

Qualitative Findings and Interpretations

Participants were asked to list the strategies they employed while completing the given task. An interpretive inductive approach (Thomas, 2006) was used to identify initial search strategies that emerged from participants listings. To reduce textual data, these were systematically grouped according to similarities (Elo & Kyngäs, 2008) and coded (Bernard & Ryan, 2010; Johnson & Christensen, 2012) into strategy descriptors. Each code was then placed in an Excel strategy descriptor list, which was ordered according to frequency (*f*) for each year level (Bernard & Ryan, 2010). To further evaluate the sophistication of strategies employed by each year level, the strategy descriptors were grouped and compared to the first and second strategy identified in the literature as part of efficient and successful Web searching.

The first strategy suggested by Neely (2006) involves the ability to clearly establish information context. This includes breaking down the research topic to establish information requirements and to identify common terminology and key concepts relevant to the task. Clearly defining and understanding the research topic aids with identifying what type of information is needed to respond to information requirements, and where best to find that information, ultimately saving time and generating more efficient search results. Thus, for the purpose of the current study, the first strategy is referred to as 'establish information context'.

The second strategy by Neely (2006) involves using the right search tools and strategies to retrieve relevant information to given context. This includes selecting the most appropriate search engines that allow access to required information and understanding the underlying search principles of selected search engines. With such understanding, the searcher is able to efficiently break down key concepts into searchable key words/phrases and effectively construct these into search strategies using appropriate language structures for the selected search engines. Also providing the searcher with the skills necessary to successfully filter and refine search strategies to yield more relevant search results. For the purpose of the current study, the second strategy will be referred to as 'search for information'.

As the current study investigated information collection only, and as students were not asked to evaluate, organise, synthesise or communicate gathered information, the third and fourth strategy identified by the literature (Johnson & Christensen, 2012; Neely, 2006; Ware, 2001) were not included in the analysis of the participant responses.

Findings

Initially, a total of 15 search strategies were extracted from participants' responses. These were grouped into 10 strategy descriptors. After being ordered according to frequency in an Excel strategy descriptor list, all descriptors mentioned less than twice across all year levels were deleted from the list. Thus, a total of 9 strategy descriptors were identified, these are listed in Table 4. The strategy descriptors were also submitted to a Chi-square analysis, however no statistically significant difference was found between the year levels.

Table 4

Overall Extracted Strategy Descriptors

Total *f*	Strategy descriptor
28	Key words/phrases
25	Filtering
13	Information requirements
9	Specific sources/websites
7	Use retreived information
5	Multiple search engines
4	Understand topic
3	Suggested links
3	Credibility

Note. f = frequency count.

The extracted strategy descriptors were then grouped according to the first and second efficient Web searching strategy identified in the literature, shown in Table 5. As the strategy descriptor *credibility* was identified as part of the third search strategy found in the literature, it was deleted.

Table 5

Overarching Search Strategies

1. Establish information context	2. Search for information
Specific information	Key words/phrases
Understand topic	Use retreived information
Specific sources/websites	Filtering
	Relevance
	Multiple search engines

Note. 1 = First strategy identified in literature. 2 = Second strategy identified in literature.

To address the research questions of the current study, each strategy descriptor employed by more than one participant within each year level was compiled and sorted from highest to lowest frequency, into a summary matrix, see Table 6. This table provides an overarching view of the underlying search strategies employed by each participant group when searching for information using the WWW to address a specific intelligence problem. Hence, addressing the principal research question; What strategies are employed by students to gather information via open sources at different stages of a tertiary Counterterrorism, Security and Intelligence course?

Table 6

Summary Matrix of Search Strategies Employed by Each Year Level in Descending Frequency

Strategies Listed as Employed by Each Cohort		
First Year	Second Year	Third Year
(2) Key words/phrases	(2) Key words/phrases	(2) Key words/phrases
(2) Filtering	(2) Filtering	(2) Filtering
(2) Multiple search engines	(1) Information requirements	(1) Information requirements
(1) Information requirements	(2) Use retrieved information	(1) Specific sources/websites
(1) Specific sources/websites	(1) Specific sources/websites	(2) Multiple search engines
	(2) Suggested links	(1) Understand topic
		(2) Use retrieved information

Note. (1) = Establish information context. (2) = Search for information.

Interpretation

The first year cohort of participants mainly listed the usage of keywords/phrases and filtering as search strategies that they employed when addressing the task. One students listed "changing the search terms to yield more results" as a search strategy, whilst other first year students listed specific keyword structures used to filter, such as + and "". However, none of the first year students acknowledged the need to understand the mechanics of the search engine to better construct their search terms, or suggested using different query language structures, such as Boolean, natural language or controlled vocabulary (Burkhardt et al., 2010; Lanning, 2012).

Although one student listed "current day data collection", and another "used key phrases that relate to the topic, like crime, police etc.". No strategy listed suggested that the participants broke down the task into information requirements (Herring, 2011) or that they looked at identifying type of information needed to respond to the requirements (Neely, 2006). Thus, the first year cohort showed least sophistication of usage of the *establish information context* and the *search for information* search strategy when addressing the task.

The second year cohort mainly listed keywords/phrases, filtering, information requirements and use retrieved information as the strategy descriptors employed. Although using different words, many of the second year participants suggested that they started their search with looking for broad and general information, and then moved forward with more narrow searches, using various techniques to filter their search in order to find relevant information. For example, one participant listed:

> First search on specific topic words in question. Add secondary words narrowing focus. Add synonymous of the main topic words. Look at the question from different perspectives and use those words. Use the information retrieved to research different paths on the topic. Multiple pages down on a search may give the information required so don't stop on page 1 of the results.

As part of filtering, second year students employed search strategies such as excluding/adding/changing/varying keywords as well as specifying that they used special query structure techniques. Some participants listed the type of information they were looking for in order to address the task, for example "how to acquire gun legally – standards/permissions", thus showing evidence of establishing information requirements. The level of sophistication of search strategies employed by the second year cohort was higher than that of the first year cohort, which becomes evident when looking at the summary matrix in Table 6.

The level of sophistication used did not differ considerably between the second and third year cohorts, although the third year cohort expressed a deeper understanding of the need to establish information context. For example, one student specifically listed "define topic" as strategy. The strategies employed to search for information were much similar to that of the second year cohort, but third year participants stated that they used specific websites/ sources, and looked for specific information, more often than the other year levels. This may indicate more elaborate consideration of what information and sources may have been useful to the specific task, and knowledge of where to find relevant information.

Although not explicitly asked of participants, the third year cohort showed evidence of considering evaluating sources when searching for information. For example, three participants listed that they stuck to, used or looked for worthy/reputable/verifiable sources as part of their strategies. This is interesting because they seem to apply this level of complexity to searching for information on the WWW, which suggest that they automatically attend to a larger number of relevant variables related to information gathering, which may reflect a level of expertise.

The comparison of the strategies employed across the three year levels of participants addressed the secondary research question, which asked; Is there evidence of progression in information gathering strategies across the three stages of the tertiary Counter Terrorism, Security and Intelligence course? From the qualitative findings it was possible to identify evidence that suggest that there is a progression between first and second year cohorts in information gathering skills. The evidence also suggests that, although less pronounced, there is some progression also between second and third year cohorts. Perhaps this is indicative of third year students refining the developed Web searching skills rather than developing new ones.

Chapter 5: Discussion and Conclusion

Summary of Study

The purpose of the current study was to gain a better understanding of how information is searched for and gathered on the WWW by potential future intelligence analysts, with the objectives to:

- Identify how students of three different progression stages of a tertiary Counter Terrorism, Security and Intelligence course go about gathering information using the WWW to address a specific intelligence problem
- Identify significant and/or meaningful differences in the approach to information gathering via the WWW between students of first, second and third year of a tertiary Counter Terrorism, Security and Intelligence course

Following the introductory Chapter 1, Chapter 2 presented the findings of existing research concerning how intelligence go about foraging information via open sources, strength and limitations of searching for information on the WWW as well as what makes a researcher an efficient and successful Web information searcher. The literature review revealed existing gaps in knowledge and thus, provided several important justifications for the significance of the current study.

Chapter 3 presented the design and methodology employed to address the objectives, principal research question, what strategies are employed by students to gather information via open sources at different stages of a tertiary Counterterrorism, Security and Intelligence course?; and secondary research question, is there evidence of progression in information gathering strategies across the three stages of the tertiary Counter Terrorism, Security and Intelligence course? Whilst certain differences in information gathering could be assessed quantitatively, other aspects, such as considerations and choices could not. Therefore, an experimental mixed method approach incorporating quantitative and qualitative methodologies was employed.

A task was (independent variable) designed where participants were asked to gather information to address a real time intelligence problem. Three dependent variables were tested at quantitatively, 'amount of information gathered', 'number of search terms used' and 'number of clicks used' in three cohorts of tertiary students. Participants were also asked to answer one open-ended question that allowed for deeper understanding of strategies and considerations employed by participants when engaging the task. As such, providing an opportunity to identify information gathering skill differences between cohorts as a function of a three-year tertiary course, which are presented in Chapter 4.

This chapter, Chapter 5, provides a discussion of the results and key findings identified in Chapter 4 in relation to the research questions of the current study. Furthermore, limitations and strengths of the current study are discussed, which leads to directions for future research as well as the conclusion of the study.

Discussion

A mixed method approach was harnessed to answer two research questions; one principal and one secondary. The principal research question was: What strategies are employed by students to gather information via open sources at different stages of a tertiary Counter Terrorism, Security and Intelligence course?

The principal research question was answered via the qualitative aspect of the current study, which revealed that students employed eight strategies overall that differed in the level of sophistication that they reflected and the frequency of their use. These eight strategies form part of the first and second efficient Web searching strategy identified in the literature (Herring, 2011; Lanning, 2012; Neely, 2006), namely 'establish information context' and 'search for information', for more details see Table 5. The first year cohort employed five strategies when gathering information on the WWW. These included the usage of key words/phrases, filtering techniques and multiple search engines as well as establishing specific information requirements of given task and gathering information directly from specific sources/websites.

The second year cohort employed six strategies, four of which were the same as the first year cohort. They did not employ the usage of multiple search engines, instead they employed the search strategies of using retrieved information to find new information and following suggested links by other sources. The third year cohort employed seven search strategies, including the five strategies employed by the first year cohort, and the usage of retrieved information employed by the second year cohort. The additional strategy employed by the third year cohort was to reflect on the topic before starting the search for more information.

The secondary research question was: Is there evidence of progression in information gathering strategies across the three stages of the tertiary Counter Terrorism, Security and Intelligence course? This question was answered both via the quantitative and the qualitative aspects. The qualitative aspect point out that there were greater elaborations of thought in the way that the third year participants tackled the task compared to both the first and second year cohorts. This suggests that students are developing and acquiring the bulk of the Web searching skills in the first year of their course. After the first year, students are not necessarily learning new skills, but rather honing and refining the skills that they have already acquired. Hence, there is evidence of progression and refinement in information gathering strategies across the three stages of the course.

The current study also found that there is a significant acquisition of skill that occurs in the first year of the course, which was identified quantitatively in the 'number of search terms used' by participants. Second year students used a significantly larger number of search terms compared with the first year participants, which also suggests that the time was used more efficiently by the second year than by the first year cohort. There were no statistically significant differences between the third and the second year, or the first and the third year cohorts.

Although the nonparametric tests found no other statistically significant differences, descriptive statistics (see Table 3) suggested trends across the three cohorts. First, the difference in both 'number of search terms used' and 'amount of information gathered' was greater between the first and second year cohorts than between the first and third year cohorts. Thus, the second year cohort used the largest number of search

terms and gathered the most amount of information, suggesting that the noticeable shift for these two variables occur during the first year of the course. This also supports the positive statistical significant correlation identified between these two variables in the quantitative analyses. Even though this was not a strong correlation, it indicated that using a greater number of search terms coincides with a greater amount of information being gathered.

Second, the difference between the second and third year cohorts was smaller significant, where the average 'amount of information gathered' and 'number of search terms used' by the third year cohort was slightly less, and slightly less dispersed than that of the second year cohort. This does not necessarily mean that no progression exists between the second and third year cohorts. It may suggest that third year students are refining rather than acquiring skills, resulting in the quality of information gathered being honed in better by the third year cohort. This suggestion was supported by the assessment of an intelligence collection expert, who evaluated and compared relevance and usability of the information gathered by each year level. The expert found that the information gathered by the third year cohort was more sophisticated and relevant to the given task than what was gathered by the second year cohort, with the first year cohort gathering the least relevant information.

Third, the mean of 'number of clicks used' was greatest in the third year cohort, with the first year cohort having the lowest. This again suggests that third year students are refining their skills rather than acquiring new skills, as they are able to look through more information faster than the first and second year students. Fourth, there was a greater disparity in the way that the second and third year cohorts searched for information compared to first year cohort, indicating that some students demonstrate greater understanding of the exploratory processes required to find information on the WWW as well as acquire and refine their skills faster than other students. Interesting to note is that the first year cohort was the most unified in the way that they were searching for information on the WWW, having the lowest disparity of the three cohorts within all three variables. Although demographic data were collected, the small number of participants achieved rendered analyses based on them meaningless.

As no literature was identified of previous studies similar to the current one, placing the findings of the current study in the broader context of the body of knowledge may be done by comparing the literature findings on efficient Web searching to that of the findings of the current study. As discussed in Chapter 2, current literature has identified a total of four overarching efficient Web searching strategies (Burkhardt et al., 2010; Herring, 2011; Johnson & Christensen, 2012; Neely, 2006), two of which were relevant to the current study. Evidence of progression across the cohorts was identified and it was beyond the current study to assess whether the students incorporated all the relevant aspects of the 'establishing information context' and/or the 'search for information' strategies. For example, no participant stated that they identified or took into account the underlying principles of the search engines (Lanning, 2012; Ware, 2001). Nor did any of the participants suggest that they looked at or identified what search engine would be most suitable to address their information requirements (Neely, 2006).

In the literature on intelligence it was identified that an intelligence analyst collecting information requires the ability to understand the breadth of information (Gill & Phythian, 2012; Moore, 2011), develop a search strategy for how and where to best forage for relevant information (Clauser, 2008; Prunckun, 2010) as well as recognise the significance of what is found when searching (Betts, 2008; Copeland, 2007). These requirements were recognised as similar to those skills suggested by the literature as necessary for a researcher to be efficient and successful when searching for information on the WWW. Hence, the current study suggests that, overall the third year student cohort that participated had developed some of the skills necessary to be successful entry-level intelligence analysts, as they evidenced understanding of breadth of information, developed search strategies and were more selective than either of the other two cohorts of the information that they gathered.

Limitations and Strengths of the Study

A number of methodological limitations as well as strengths have been recognised within the current study, which present opportunities for future research. The main limitations stem from time constraints and the limited sample size. Data collection

involved a relatively small number of participants (N = 40), owed at large to the self-selected nature of participation as well as the constrained number of students enrolled in each year level of the course. This saw an insufficient number of participants for statistical analysis of demographic data and for the, originally suggested, parametric statistics. Whilst nonparametric analyses were appropriately employed, the current study suggests that future research should consider recruiting larger samples.

Pairwise effect size calculations were conducted, which revealed that the participant numbers were insufficient to reliably establish statistically significant differences between some year levels. In order to achieve power of .8, the desired number of participants for the variable 'amount of information gathered' was calculated to be 12 between the first and second year cohorts, and 78485 between the second year and third year cohorts. The desired number of participants between the second and third year cohorts was calculated to be 78485 for the variable 'number of search terms used'. For the variable 'number of clicks used', the desired number of participants was calculated to be 28 between first and second year cohorts, 17 between first and third year cohorts and 268 between second and third year cohorts. Hence, future studies informed by the effect size calculations of the current study and utilising similar variables should take this into consideration. The findings for the other comparisons should be reliable as the sample sizes calculated for these were identified as adequate.

Additionally, even though intelligence analysts are recruited from within various tertiary disciplines, within the constraints of an Honours project it was beyond the scope to do a cross-disciplinary analysis or comparison between different tertiary disciplines. As the sample population represented those who have a pre-expressed interest in intelligence, and are the most likely to become entry-level intelligence analysts, they were identified as appropriate for the current study, however, it may be suggested that future research could allow for such comparisons. Further research into the current area of study is also suggested to test the conclusions and recommendations drawn from the data.

A number of strengths have been acknowledged in the current study. First, it is the first study of its kind as no other prior to this has attempted to investigate the potential progression of Web information foraging skills among up coming intelligence

analysts. Second, the study used a multi-method approach that allowed corroboration of findings from both quantitative and qualitative perspectives. Third, the study utilised an actual real time intelligence problem as the experimental task, thus enhancing external validity. Fourth, the study used standardised procedures, which minimises experimenter bias as well as demand characteristics across the experimental groups. Fifth, careful consideration was given to the utilisation of statistical procedures that matched level of measurement, distribution characteristics and sample sizes so to produce results responsibly. Sixth, the settings and procedures used were similar to those encountered by intelligence analysts in the field, hence ecological validity was addressed.

Directions for Future Research

Whilst the current study served to narrow the gap in knowledge relating to how intelligence analysts forage for information using open sources such as the WWW, there is still a need for more research into this area. Given the findings of the current study, there are several avenues open for further research. As the intelligence domain recruits new entry-level analysts from various tertiary disciplines, further research aimed at identifying how students within different tertiary disciplines search for and gather information on the Web would be beneficial. In addition, further research into how tertiary students go about evaluating information and sources found on the WWW, as well as how they synthesise gathered information in order to produce an intelligence product would be beneficial as it provides a holistic understanding of the skill set of potential future intelligence analysts. Finally, in depth qualitative as well as quantitative investigations with experienced analysts could inform tertiary courses toward facilitating the development of up coming analysts' skills.

Conclusion

Whilst further research and replication of the current study are recommended considering the limitations mentioned above. The current study found that eight strategies overall were employed by students to gather information via the WWW at different stages of the tertiary Counter Terrorism, Security and Intelligence course. These varied in the frequency of their use as well as the level of sophistication that they

reflected. The findings of the current study also suggest that the greatest Web searching skill acquisition occurs in the first year of the course, which is followed by further refinements of those skills throughout the second and third year of the course. This has been reflected in the statistically significant difference that was found in the number of search terms used between first and second year participants, and in the qualitative differences that reflected growing expertise and greater elaboration and considerations in the strategies employed by the third year, compared to the first and second year of the course.

References

AFCEA Intellgence Committee. (2005). Making analysis relevant: It's more than connecting the dots *White Paper*. Fairfax, VA: Armed Forces Communications and Electronics Assiciation.

Agosto, D. E. (2002). Bounded rationality and satisficing in young people's Web-based decision making. *Journal of the American Society for Information Science and Technology, 53*(1), 16-27. doi: 10.1002/asi.10024

Alexander, J. E., & Tate, M. A. (1999). *Web wisdom: how to evaluate and create information quality on the Web*. Mahwah, N.J: Lawrence Erlbaum Associates.

Appel, E. J. (2011). *Internet Searches for Vetting, Investigations, and Open-Source Intelligence*. Hoboken: CRC Press.

Australia Security Intelligence Organisation [ASIO]. (2013). Intelligence Professional Retrieved 09 May, 2013, from http://www.asio.gov.au/Careers/ASIO-Careers/Intelligence-Professionals.html

Bar-Ilan, J. (2007). Google Bombing from a Time Perspective. *Journal of Computer-Mediated Communication, 12*(3), 910-938. doi: 10.1111/j.1083-6101.2007.00356.x

Bean, H. P. D. (2011). *No More Secrets: Open Source Information and the Reshaping of U.S. Intelligence*. Santa Barbara: ABC-CLIO.

Bernard, H. R., & Ryan, G. W. (2010). *Analyzing qualitative data: systematic approaches*. Thousand Oaks, Calif: Sage.

Betts, R. K. (2008). Two faces of intelligence failure: September 11 and Iraq's missing WMD. *Political Science Quarterly, 122*(4), 585-606.

Biddix, J. P., Chung, C. J., & Park, H. W. (2011). Convenience or credibility? A study of college student online research behaviors. *The Internet and Higher Education, 14*(3), 175-182. doi: 10.1016/j.iheduc.2011.01.003

Bilal, D. (2000). Children's Use of the Yahooligans! Web Search Engine: I. Cognitive, Physical, and Affective Behaviors on Fact-based Search Tasks. *Journal of the American Society for Information Science, 51*(7), 646-665.

Bilal, D. (2001). Children's use of the Yahooligans! Web search engine: II. Cognitive and physical behaviors on research tasks. *Journal of the American Society for Information Science and Technology, 52*(2), 118-136. doi: 10.1002/1097-4571(2000)9999:9999<::aid-asi1038>3.3.co;2-i

Bilal, D. (2002). Children's use of the Yahooligans! Web search engine. III. Cognitive and physical behaviors on fully self-generated search tasks. *Journal of the American Society for Information Science and Technology, 53*(13), 1170-1183. doi: 10.1002/asi.10145

Bilal, D., & Bachir, I. (2007). Children's interaction with cross-cultural and multilingual digital libraries. II. Information seeking, success, and affective experience. *43*(1), 65-80. Retrieved from doi:10.1016/j.ipm.2006.05.008

Bodnar, J. W. (2003). Warning analysis for the information age: Rethinking the intelligence process R. Swenson (Ed.) (pp. 193). Retrieved from http://www.higginsctc.org/intelligence/warning_analysis.pdf

Bowler, L., Large, A., & Rejskind, G. (2001). Primary School Students, Information Literacy and the Web. *Education for Information, 19*(3), 201-223.

Braasch, J. L. G., Lawless, K. A., Goldman, S. R., Manning, F. H., Gomez, K. W., & MacLeod, S. M. (2009). Evaluating Search Results: An Empirical Analysis of Middle School Students' Use of Source Attributes to Select Useful Sources. *Journal of Educational Computing Research, 41*(1), 63-82.

Bradbury, D. (2011). In plain view: open source intelligence. *Computer Fraud & Security, 2011*(4), 5-9. doi: 10.1016/s1361-3723(11)70039-2

Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust Matters: Examining the Role of Source Evaluation in Students' Construction of Meaning within and across Multiple Texts. *Reading Research Quarterly, 44*(1), 6-28.

Brem, S. K., Russell, J., & Weems, L. (2001). Science on the Web: student evaluations of scientific arguments. *Discourse Processes, 32*(2/3), 191-213. doi: 10.1207/s15326950dp3202&3_06

Britt, M. A., & Aglinskas, C. (2002). Improving Students' Ability to Identify and Use Source Information. *Cognition and Instruction, 20*(4), 485-522. doi: 10.1207/s1532690xci2004_2

Burkhardt, J. M., MacDonald, M. C., & Rathemacher, A. J. (2010). *Teaching information literacy: 50 standards-based exercises for college students*. Chicago: American Library Association.

Camacho, D., Aler, R., Borrajo, D., & Molina, J. M. (2006). Multi-agent plan based information gathering. *Applied Intelligence, 25*(1), 59-71. doi: 10.1007/s10489-006-8866-z

Canton, B. (2008). The Active Management of Uncertainty. *International Journal of Intelligence and CounterIntelligence, 21*(3), 487-518. doi: 10.1080/08850600802046939

Choo, C. W., Turnbull, D., & Detlor, B. (2000). *Web work: information seeking and knowledge work on the World Wide Web* (Vol. 1.). Dordrecht: Kluwer Academic.

Clark, R. M. (2010). *Intelligence analysis: a target-centric approach* (3rd ed.). Washington, D.C: CQ Press.

Clauser, J. (2008). *An introduction to intelligence research and analysis* (Vol. 3.). Lanham, Md: Scarecrow Press.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. New York: Routledge.

Coiro, J. (2003). Reading Comprehension on the Internet: Expanding Our Understanding of Reading Comprehension To Encompass New Literacies. *Reading Teacher, 56*(5), 458-464.

Coiro, J. (2005). Making Sense of Online Text. *Educational Leadership, 63*(2), 30.

Coiro, J. (2009). Rethinking online reading assessment. *Educational Leadership, 66*(6), 59.

Coiro, J. (2011). Predicting Reading Comprehension on the Internet: Contributions of Offline Reading Skills, Online Reading Skills, and Prior Knowledge. *Journal of Literacy Research, 43*(4), 352-392. doi: 10.1177/1086296x11421979

Coiro, J., & Dobler, E. (2007). Exploring the Online Reading Comprehension Strategies Used by Sixth-Grade Skilled Readers to Search for and Locate Information on the Internet. *Reading Research Quarterly, 42*(2), 214-257.

Cooper, J. R. (2005). *Curing analytic pathologies: Pathways to improved intelligence analysis*. Washington: Central Intelligence Agency, Center for the Study of Intelligence. Retrieved from https://http://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/curing-analytic-pathologies-pathways-to-improved-intelligence-analysis-1/analytic_pathologies_report.pdf.

Copeland, T. E. (2007). *Fool me twice: intelligence failure and mass casualty terrorism*. Leiden: Martinus Nijhoff.

Creswell, J. W. (2009). *Research design: qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, Calif: Sage Publications.

de Kunder, M. (2012). The size of the World Wide Web (the Internet). *Daily estimated size of the World Wide Web* Retrieved 17 Sep, 2012, from http://www.worldwidewebsize.com/

Department of Defence Intelligence and Security [DIO]. (2013). Frequently Asked Questions: What degree should i have? Retrieved 09 May, 2013, from http://www.defence.gov.au/dio/faq.shtml

Downes, P. K. (2007). Introduction to the World Wide Web. *BDJ, 202*(9), 535-540. doi: 10.1038/bdj.2007.371

Doyle, T., & Hammond, J. L. (2006). Net cred: evaluating the internet as a research source. *Reference Services Review, 34*(1), 56-70. doi: 10.1108/00907320610648761

Dresang, E. T. (2005). The Information-Seeking Behavior of Youth in the Digital Environment. *Library Trends, 54*(2), 178.

Edith Cowan University [ECU]. (2012). Course Information: Y14 Bachelor of counter terrorism security and intelligence. *Handbook 2012* Retrieved 17 Sep, 2012, from http://handbook.ecu.edu.au/CourseStructure.asp?disyear=2012&CID=1406&USID=0&UCID=0&UID=0&Ver=0.0302&HB=HB&SC=UG

Eliopoulos, D., & Gotlieb, C. (2003). Evaluating Web search results rankings. *Online, 27*(2), 42.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of advanced nursing, 62*(1), 107-115. doi: 10.1111/j.1365-2648.2007.04569.x

Federal Bureau of Investigation [FBI]. (n.d.). IA Qualification Requirements Retrieved 09 May, 2013, from https://http://www.fbijobs.gov/1212.asp

Fidel, R., Davies, R. K., Douglass, M. H., Holder, J. K., Hopkins, C. J., Kushner, E. J., . . . Toney, C. D. (1999). A Visit to the Information Mall: Web Searching Behavior of High School Students. *Journal of the American Society for Information Science, 50*(1), 24-37.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education*. Boston: McGraw-Hill.

Gazit, S. (1980). Estimates and Fortune-Telling in Intelligence Work. *International Security, 4*(4), 36-56.

George, R. Z. (2004). Fixing the Problem of Analytical Mind-Sets: Alternative Analysis. *International Journal of Intelligence and CounterIntelligence, 17*(3), 385-404. doi: 10.1080/08850600490446727

Gersh, J., Lewis, B., Montemayor, J., Piatko, C., & Turner, R. (2006). Supporting insight-based information exporation in intelligence analysis. *Communications of the ACM, 49*(4), 63-68.

Gibson, S. (2004). Open source intelligence. *The RUSI Journal, 149*(1), 16-22. doi: 10.1080/03071840408522977

Gill, P., & Phythian, M. (2012). *Intelligence in an insecure world* (2nd ed.). Malden, MA: Polity Press.

Glassman, M., & Kang, M. J. (2012). Intelligence in the internet age: The emergence and evolution of Open Source Intelligence. *Computers in Human Behavior, 28*(2), 673. doi: 10.1016/j.chb.2011.11.014

Grabo, C. M. (2004). *Anticipating surprise: analysis for strategic warning*. Lanham, Maryland: University Press of America.

Gross, M., & Latham, D. (2007). Attaining information literacy: An investigation of the relationship between skill level, self-estimates of skill, and library anxiety. *Library and Information Science Research, 29*(3), 332-353. doi: 10.1016/j.lisr.2007.04.012

Gulli, A., & Signorini, A. (2005). *The indexable web is more than 11.5 billion pages.* Paper presented at the 14th International World Wide Web Conference, Chiba, Japan.

Herring, J. E. (2011). *Improving students' web use and information literacy: a guide for teachers and teacher librarians*. London: Facet.

Heuer, R. J. (2005). Limits of intelligence analysis. *Orbis, 49*(1), 75-94.

Heuer, R. J., & Pherson, R. H. (2011). *Structured analytic techniques for intelligence analysis*. Washington, DC: CQ Press.

Hirsh, S. G. (1999). Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science, 50*(14), 1265-1283.

Hoffman, J. L., Wu, H.-K., Krajcik, J. S., & Soloway, E. (2003). The nature of middle school learners' science content understandings with the use of online resources. *Journal of Research in Science Teaching, 40*(3), 323-346. doi: 10.1002/tea.10079

Howell, D. C. (2008). *Fundamental statistics for the behavioral sciences*. Belmont, CA: Thomson/Wadsworth.

Hulnick, A. S. (2006). What's wrong with the intelligence cycle. *Intelligence and National Security, 21*(6), 959-979. doi: 10.1080/02684520601046291

Johnson, B., & Christensen, L. B. (2012). *Educational research: quantitative, qualitative, and mixed approaches*. Thousand Oaks, Calif: SAGE Publications.

Jones, S., Johnson-Yale, C., Millermaier, S., & Pérez, F. S. (2008). Academic work, the Internet and U.S. college students. *The Internet and Higher Education, 11*(3-4), 165-177. doi: 10.1016/j.iheduc.2008.07.001

Judd, T., & Kennedy, G. (2010). A five-year study of on-campus Internet use by undergraduate biomedical students. *Computers & Education, 55*(4), 1564-1571. doi: 10.1016/j.compedu.2010.06.022

Kim, & Sin. (2011). Selecting quality sources: Bridging the gap between the perception and use of information sources. *Journal of Information Science, 37*(2), 178-188. doi: 10.1177/0165551511400958

Lanning, S. (2012). *Concise Guide to Information Literacy*. Santa Barbara: ABC-CLIO.

Large, A., & Beheshti, J. (2000). The Web as a Classroom Resource: Reactions from the Users. *Journal of the American Society for Information Science, 51*(12), 1069-1080. doi: 10.1002/1097-4571(2000)9999:9999<::aid-asi1017>3.0.co;2-w

Large, A., Beheshti, J., & Breuleux, A. (1998). Information seeking in a multimedia environment by primary school students. *Library and Information Science Research, 20*(4), 343-376. doi: 10.1016/s0740-8188(98)90027-5

Lefebvre, S. (2004). A Look at Intelligence Analysis. *International Journal of Intelligence and CounterIntelligence, 17*(2), 231-264. doi: 10.1080/08850600490274908

Lewandowski, D. (2008). A three-year study on the freshness of web search engine databases. *Journal of Information Science, 34*(6), 817-831. doi: 10.1177/0165551508089396

Lorenzen, M. (2002). The Land of Confusion? High School Students and Their Use of the World Wide Web for Research. *Research Strategies, 18*(2), 151-163.

Maybee, C. (2007). Understanding our student learners. *Reference Services Review, 35*(3), 452-462. doi: 10.1108/00907320710774319

Mercado, S. C. (2004). Sailing the Sea of OSINT in the Information Age: A Venerable Source in a New Era. *Studies in Intelligence, 48*(3), 45-55.

Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student Web use, perceptions of information credibility, and verification behavior. *Computers & Education, 41*(3), 271-290. doi: 10.1016/s0360-1315(03)00049-6

Moore. (2011). Sensemaking: A structure for an intelligence revolution Clift Series on the Intelligence Profession, (pp. 233).

Moore, D. T., Krizan, L., & Moore, E. J. (2005). Evaluating Intelligence: A Competency-Based Model. *International Journal of Intelligence and CounterIntelligence, 18*(2), 204-220. doi: 10.1080/08850600590911945

Nachmias, R., & Gilad, A. (2002). Needle in a hyperstack: Searching for information on the World Wide Web. *Journal of Research on Technology in Education, 34*(4), 475.

Neely, T. Y. (2006). *Information literacy assessment: standards-based tools and assignments*. Chicago: American Library Association.

Olcott, A. (2012). *Open Source Intelligence in a Networked World*. London: Continuum International Publishing.

Plano Clark, V. L., & Creswell, J. W. (2008). *The mixed methods reader*. Thousand Oaks, Calif: Sage Publications.

Prunckun, H. (2010). *Handbook of scientific methods of inquiry for intelligence analysis* (Vol. 11). Lanham: Scarecrow Press.

Quarmby, N., & Young, L. J. (2010). *Managing intelligence: the art of influence*. Annandale, N.S.W: Federation Press.

Richards, J. (2010). *The art and science of intelligence analysis*. Oxford: Oxford University Press.

Rodgers, R. S. (2006). Improving Analysis: Dealing with Information Processing Errors. *International Journal of Intelligence and CounterIntelligence, 19*(4), 622-641. doi: 10.1080/08850600600829858

Scott, L., & Jackson, P. (2004). The study of intelligence in theory and practice. *Intelligence & National Security, 19*(2), 139-169. doi: 10.1080/0268452042000302930

Sims, J. E., & Gerber, B. L. (2005). *Transforming U.s. Intelligence*. Georgetown: Georgtown University Press.

Spink, A., Jansen, B. J., Kathuria, V., & Koshman, S. (2006). Overlap among major web search engines. *Internet Research, 16*(4), 419-426. doi: 10.1108/10662240610690034

Stamp, M. (2011). *Information security: principles and practice*. Hoboken: John Wiley & Sons.

Steele, R. D. (2007). Open source intelligence. In L. K. Johnson (Ed.), *Handbook of intelligence studies* (pp. 129-147). Hoboken: Taylor & Francis.

Steele, R. D., & Press, O. S. S. I. (2006). *Information operations: all information, all languages, all the time*. Oakton, Va: OSS, Inc.

Stephen, M. (2012). Is Intelligence Analysis an Art or a Science? *The International Journal of Intelligence and Counter Intelligence, 25*(3), 529.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Pearson/Allyn & Bacon.

Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation, 27*(2), 237-246. doi: 10.1177/1098214005283748

Thompson, C. (2003). Information illiterate or lazy: How college students use the web for research. *Portal: Libraries and the Academy, 3*(2), 259-268. doi: 10.1353/pla.2003.0047

Tu, Y.-W., Shih, M., & Tsai, C.-C. (2008). Eighth graders' web searching strategies and outcomes: The role of task types, web experiences and epistemological beliefs. *Computers & Education, 51*(3), 1142-1153. doi: 10.1016/j.compedu.2007.11.003

Uyar, A. (2009). Investigation of the accuracy of search engine hit counts. *Journal of Information Science, 35*(4), 469-480. doi: 10.1177/0165551509103598

Walsh, P. F. (2011). *Intelligence and intelligence analysis*. Abingdon, Oxon: Routledge.

Ware, M. (2001). Evaluating information on the Internet. *School Libraries Worldwide, 7*(2), 39.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source Evaluation, Comprehension, and Learning in Internet Science Inquiry Tasks. *American Educational Research Journal, 46*(4), 1060-1106. doi: 10.3102/0002831209333183

Wolfe, C. R. (2001). *Learning and teaching on the World Wide Web*. London: Academic.

Appendix A: Information Letter

Dear Participant,

My name is Teresa Cunow and I would like to invite you to participate in a study that forms part of my course requirements as a Security Science Honours student. This study aims to investigate information collection skills using open sources. This is an important and topical issue for the intelligence domain, academia, and the security industry. As the intelligence profession mostly recruits university graduates as analysts, I invite you to participate in the study. Your participation in the research would be most appreciated. This study has gained ethics approval from the Faculty of Computing Health and Science at Edith Cowan University.

Participation in this study is voluntary and you may withdraw your participation at any stage without penalty. The study involves gathering information through open sources about a specific topic and would take about one hour to complete. Only demographic information will be asked for. Confidentiality is assured and no personal identifying information will be collected. The experiment will be conducted at computer labs on Joondalup campus during assigned tutorial times. As these are typical computer labs, because of occupational health and safety regulations, all participants must wear closed shoes when entering the lab.

If you would like to participate in this study or require further information about this project, please contact me, **Teresa Kasprzyk Cunow**, on email. **skasprzy@our.ecu.edu.au,** or my Supervisor, **Mr. Jeff Corkill** (ph. **6304 5544**, email. **j.corkill@ecu.edu.au**). If you have any concerns about the project or would like to talk to an independent person, you may contact the Honours Co-ordinator, Edith Cowan University – Dr Martin Masek (ph **9370 6410**, email. **m.masek@ecu.edu.au**).

Appendix B: Consent Form

A Study on the Effects of Tertiary Education on Open Source Information Gathering Skills

I _____ (the participant) have read the information sheet provided with this consent form and any questions I have asked have been answered to my satisfaction.

I agree to participate in this study, realising I may withdraw at any time without penalty.

I agree that the research data gathered for this study be used to complete a publishable research report provided that I am not identifiable.

I understand that by participating, or opting not to participate in this study I will not be disadvantaged in any way in my course of study.

I understand that I will be asked to answer a number of demographic questions about myself such as my age, sex as well as years of eduaction and I agree to such. I also understand that I will not be asked to provide identifying information such as my name and/or student number.

If you would like to have a summary of the results sent to you in due course via your ECU student email, please tick the box below:

☐

Participant's signature _____ Date _____

Researcher's signature _____ Date _____

Appendix C: Task Instructions

Task:

Your task is to gather information that you see relevant about gun crime in Western Australia.

What you are required to do:

Search and gather information about the given topic using open sources for a period of 30 minutes. Please copy each search term you use, and the number of results each search term generates, and paste into the provided word document. Please also copy the URL and the element of the page that is relevant of the information you want to gather and paste into the assigned field in the provided word document. Once the 30 minutes are up, please save the new information entered into the provided word document.

How to copy, paste and save information:

If you normally use a different method to copy and paste than suggested, you may continue to use it here. However, if you are unfamiliar with how to copy, paste and save information, please follow the provided instructions.

Mark the appropriate information by pressing down the left click on the mouse whilst dragging the curser over the search terms or text you want to copy. Once appropriate information is highlighted you can either right click on the mouse and choose copy in the drop down list, or follow the instructions provided below regarding keyboard instructions. Paste the copied information into the assigned column in the Word document provided by either placing the cursor in that field and right clicking (here chose “paste” in the drop down list) or by following the keyboard instructions below. Save the new information by either clicking on the floppy disk icon on the word document bar or by following below keyboard instructions.

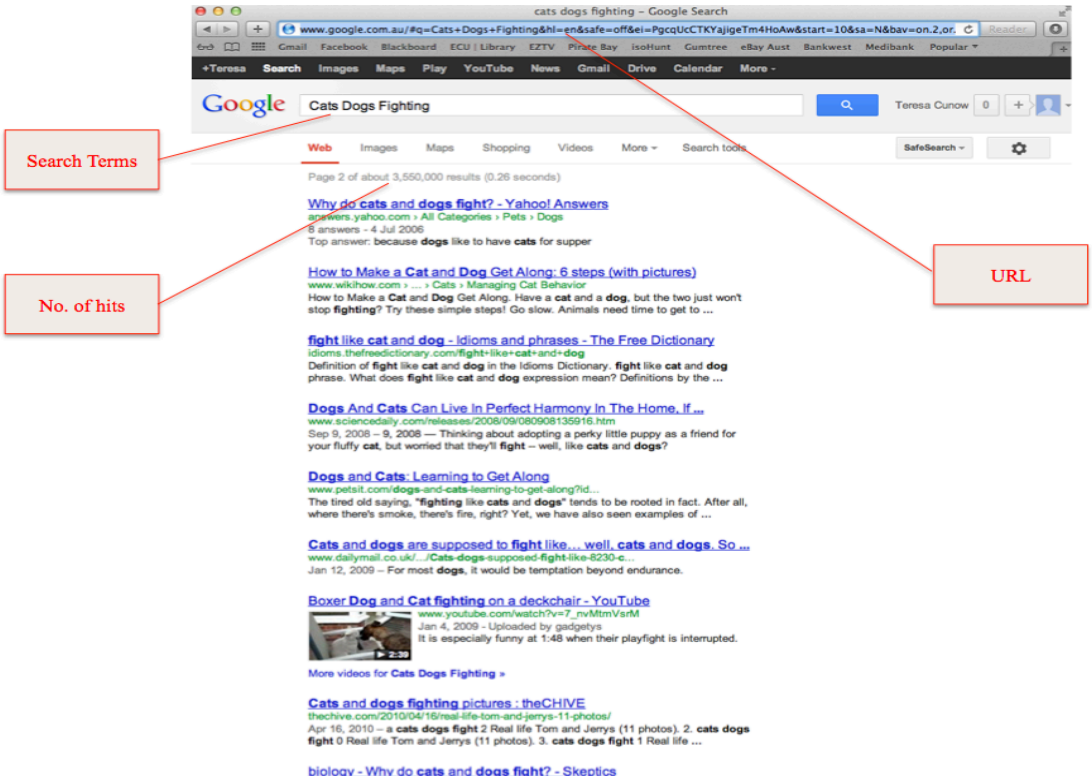
Keyboard shortcuts:

CTRL + C = copy

CTRL + V = paste

CTRL + S = save

Appendix D: Example Word Data Collection Sheet



Search Terms:	Number of hits returned by search terms:	Information (URL and the element of the page that is relevant):
Example: Cats Dogs Fighting	**Example:** 3,550,000	**Example:** http://www.sciencedaily.com/releases/2008/09/080908135916.htm New research at Tel Aviv University, the first of its kind in the world, has found a new recipe for success. According to the study, if the cat is adopted before the dog and if they are introduced when still young (less than 6 months for kittens, a year for dogs), there is a high probability that your two pets will get along swimmingly.
		Example: http://www.petsit.com/dogs-and-cats-learning-to-get-along?id=332213 The most recent statistics from the American Veterinary Medical Association report that 44 percent of U.S. pet owners have multiple-pet households, and the most common combinations include dogs and cats in the mix. Since this combination is the most likely, it's important to know how to handle the situation IF the family pets aren't playing well together.

Appendix E: Demographic Information

1. What is your sex?

☐ Female ☐ Male

2. What was your age in years at your last birthday? _____

3. Is English your first language?

☐ Yes ☐ No

4. How many years of formal education have you had overall, regardless of completion?

_____

5. Which year in the Bachelor of Counterterrorism, Security and Intelligence are you currently enrolled in?

1st	2nd	3rd
☐	☐	☐

6. Do you have any previous work experience that is related to information gathering and/or information analysis?

☐ Yes ☐ No

If yes, how many years? _____

Appendix F: Open Ended Question

In dot form, please list the strategies you employed to achieve the given task: