

PERSONAL INFORMATION SEARCHING AND IDENTIFICATION IN
PERSONAL COMPUTER

NOORASHEILA BINTI JAAFAR

A project report submitted in partial fulfillment of the
requirements for the award of the degree of
Master of Computer Science (Information Security)

Faculty of Computing
Universiti Teknologi Malaysia

AUGUST 2013

This project report is dedicated to my family and friends for their endless support,
understanding and encouragement.

ACKNOWLEDGEMENT

Alhamdulillah, thank you Allah s.w.t for giving me the strength to finally finish my study. I would like to express my heartfelt gratitude to my supervisor, Dr. Shukor Abd Razak for his constant support and understanding. Dr. Shukor has helped me greatly in order to finish this project. I have learned a lot from him and forever thankful to have him as my mentor and supervisor.

I also want to thank my friends for their encouragement and help during my time here in UTM. They are with me when I am in need; they are my true friend indeed. Not forgetting my gratitude to Universiti Teknologi Malaysia and all its staff members for providing me with a comfortable environment and facilities to complete my study.

Finally, I am grateful to my family. No words can describe how lucky I am to have them. Mom and Dad, this is for you.

ABSTRACT

Personal computer (PC) is a microprocessor designed for an individual user. Over few decades, personal computer has evolved into high power machine which plays huge role in our daily lives, both for work and play. We use personal computer for communication, internet browsing, online transaction and social networking. We create files and emails and during all these activities, our PC collects information. After years of use, in all these exchange a notable amount of personal information and sensitive data are stored within our PC. Personal information is detail used to distinguish and identify an individual. The personal information stored within PC is not used regularly by user and is in fact can be considered as a liability. Disclosure of personal information exposed users to the risk of being a victim to hackers. Hence, user should regularly check the information gathered within PC and clean any unnecessary personal information. This research examined on categories of personal information commonly found on personal computer. A technique was proposed which helps user to search the existence and identify the location of personal information in the PC. The collected data from the technique is then analyzed. The proposed technique was also tested and some future works are suggested at the end of this study.

ABSTRAK

Komputer peribadi merupakan mikropemproses yang direka khas untuk kegunaan seorang individu. Sejak beberapa dekad, komputer peribadi telah berkembang menjadi mesin berkuasa tinggi yang memainkan peranan penting dalam kehidupan seharian manusia sama ada untuk berkerja atau berhibur. Kita menggunakan komputer peribadi untuk berkomunikasi, melayari internet dan juga rangkaian sosial. Kita mereka fail dan emel. Semasa melakukan semua aktiviti ini, komputer peribadi mengumpul maklumat. Selepas bertahun digunakan, sejumlah maklumat peribadi dan sensitif yang ketara telah disimpan dalam komputer peribadi. Maklumat peribadi adalah rekod yang digunakan untuk membezakan dan mengenali seseorang individu. Maklumat peribadi ini tidak digunakan dengan kerap oleh pengguna komputer malah boleh dianggap sebagai liabiliti. Pendedahan maklumat peribadi menyebabkan pengguna terdedah kepada risiko menjadi mangsa penggadam komputer dengan niat jahat. Oleh itu, pengguna perlu kerap memeriksa maklumat yang terdapat dalam komputer peribadi dan membuang segala maklumat peribadi yang tidak diperlukan. Penyelidikan ini akan memeriksa maklumat peribadi yang selalunya tersimpan dalam komputer peribadi. Satu teknik dicadangkan untuk membantu pengguna mencari kewujudan dan mengenalpasti lokasi maklumat peribadi dalam komputer peribadi pengguna. Data yang dikumpulkan dianalisa dan teknik yang dicadangkan juga diuji. Kajian dimasa hadapan telah dicadangkan diakhir kajian ini.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xi
	LIST OF FIGURES	xii
1	INTRODUCTION	
	1.0 Introduction	1
	1.1 Problem Background	2
	1.2 Problem Statement	4
	1.3 Project Purpose	4
	1.4 Project Objective	4
	1.5 Project Scope	5
	1.6 Significant of Project	5
	1.7 Project Organization	6
2	LITERATURE REVIEW	
	2.1 Introduction	7
	2.2 Computer	8
	2.2.1 Computer Types	8
	2.3 A Brief History of Personal Computer	9
	2.4 A Brief History of Laptops	11
	2.4.1 The Laptop Revolution	12

2.5	Personal Information	15
2.5.1	Personal Information and Privacy Concern	16
2.5.2	Personal Identifiable Information	17
2.6	Confidentiality Impact Level	19
2.7	Classification of Personal Information	19
2.8	Proposed Personal Information Classification	24
2.9	Information Searching: Windows 7 Search	25
2.10	Notable Case regarding with Personal Computer	27
2.11	Summary	28
3	METHODOLOGY	
3.1	Introduction	29
3.2	Research Framework	30
3.2.1	Phase 1: Personal Information Classification	33
3.2.2	Phase 2: The Proposed Technique	34
3.2.3	Phase 3: Analysis	36
3.3	Summary	
4	DESIGN AND IMPLEMENTATION	
4.1	Introduction	37
4.2	Overview of Proposed Technique	37
4.2.1	What the crawler scan	38
4.2	The Regular Expression	39
4.3	Crawler Implementation	44
4.4	The Crawling Process	44
4.5	Summary	47
5	FINDINGS AND DISCUSSION	
5.1	Introduction	48
5.2	Proposed Personal Information Classification	49
5.3	The Log report	51
5.4	Analysis of Findings	54
5.4.1	Participants Feedback	56

5.5	Summary	59
6	CONCLUSION AND RECOMMENDATION	
6.1	Introduction	60
6.2	Research Findings	60
6.3	Research Contributions	62
6.4	Future Works	62
6.5	Conclusion	64
	REFERENCES	65
	APPENDIX A	71

LIST OF TABLE

TITLE	PAGE
Table 2.1: Different Types of Personal Computer	10
Table 2.2: A Look at laptop History	13
Table 2.3: Examples of PII	18
Table 3.1: Overview of research plan	30
Table 4.1: Regular Expression	40
Table 4.2: Example of Match Content of regexep	41
Table 4.3: Keyword Match	41
Table 5.1: Personal Information Classification	49

LIST OF FIGURES

TITLE	PAGE
Figure 1.1: Project Organization	6
Figure 2.1: Different Power Types of Computers	8
Figure 2.2: Search Box	26
Figure 3.1: Research Framework	32
Figure 4.1: Flowchart of the Crawler Design	46
Figure 5.1: Log Report	51
Figure 5.2: Excerpt 1 of Log view	52
Figure 5.3: Excerpt 2 of Log view	53
Figure 5.4: The amount of personal information match collected	55
Figure 5.5: Participants Feedback Questions	56
Figure 5.6: Participants Responds	57
Figure 6.1: Expected OCR Implementation	63

CHAPTER 1

INTRODUCTION

1.0 Introduction

Over few decade personal computer have evolved into high power machines which plays huge role in our daily lives, both for work and play. Almost all the time we interact with computers. Be it a personal computer, a laptop or notebook, desktop computer or even a tablet PC, almost everyone has at least one of these items. What exactly is a personal computer? We used personal computer every day, but we do not really delve into the definition of a personal computer. Generally speaking, a personal computer or PC is a microcomputer designed for an individual user. Previously, many years ago, computer was designed for companies and was extremely expensive, hence only affordable by these particular companies. Back then, terminals will be attached for multiple users, to one single large computer which hold resources which then shared among all users.

Today, personal computers are a lifestyle. Most Malaysians have at least one personal computer in their home, either for work or for entertainment, and some have one computer for each of their family members. Computers make life easier. With internet advancement today, people can do almost everything at

home. We do need to go to shopping mall for grocery shopping or buys household appliances. We do not need to go to the cinema to watch our favourite movies. Everything can be done through online, comfortably resting at home.

Computer may also be some form of diary. We can share stories and information with friends from around the world through social networking, therefore, computer holds variety of information of ourselves and anyone who interact with that particular computer. However, there are also disadvantages of relying to personal computer on everyday work that unsurprisingly, not many noticed of. Each time we browse the net, we do social networking, we create files, emails and many more activities our personal computer collects information. After a few years, months or even days, in all these exchange, a notable amount of personal information and sensitive data are stored within the personal computer (Bozidar Spirovski, 2010). Most of these information are not used regularly by user, in fact can be considered as a liability, hence, it is a good practice to check the information gathered and clean any unnecessary personal information.

1.1 Problem Background

Nowadays, there are many available applications to recover deleted or lost files. However, it is very hard to find application which helps user to locate and search files. As mentioned above, it is a very good practice to check the information and personal data gathered by your personal computer from time to time and clean and delete any unnecessary personal information. Unfortunately, not all users are technical; hence they need help to scans locations within their personal computer for sensitive data. Our personal computer collects information even without user intervention, for example the cc function in emails. Our browsing history and cookies contains personal information which can become a liability if it falls into wrong hands such as hackers or people with malicious

intent. Moreover, ordinary users are not exposed enough to the issue regarding privacy concerns. Hence, a technique which can help user to search for personal information within their personal computer is significantly helpful. Protecting one's personal information is important in order to maintain one's information security.

Personal information according to (McCallister et al. 2010) can be used to uniquely identify, contact or locate a single person. Personal information can also combined with other resources to uniquely single out an individual. Unauthorized disclosure of personal information could result a serious adverse effect to the particular user and cause damage to reputation and also invasion of privacy.

On the other hand, personal information within a personal computer might become evidence in legal and criminal investigations. A suspect's personal computer might hold crucial personal information which is a significant evidence to prosecute the suspect. In any forensic investigation, gathering every single significant evidence and information quickly and accurately is very crucial in order to solve the crime investigated properly. A technique which search personal information in personal computer swiftly and accurately is very helpful in crime investigations. Through personal information, law enforcements may establish motives, personal relationship and prove or disprove alibis of suspected individual (Lynch and Ellickson, 2010).

1.2 Problem Statement

Over years or months, our personal computer gathered significant amount of personal information. These personal data is not actually much of use to user daily works. Therefore, it is good practice to clean all the unnecessary personal information every now and then to avoid risks being a prey of hackers with malicious intent. A technique which can accurately locate and find personal information within a personal computer is very helpful for this purpose.

1.3 Project Purpose

The purpose of this project is to propose a technique when launched within a personal computer will search and identify all personal information stored in the particular personal computer.

1.4 Project Objective

In order to ensure that this project could be completed successfully, the following objectives have been identified;

- i) To propose categories of personal information classification within personal computers.
- ii) To propose a new technique for personal information search and identify within a personal computer.

- iii) To test and validate the proposed technique for searching and identify personal information.

1.5 Project Scope

The following are the scopes identified to be followed during the process of this project:

- i) The technique is designed for personal computer only.
- ii) The technique will be tested on existing personal computer.
- iii) The technique is for searching and identifying personal information only.

1.6 Significant of Project

This project is prepared to produce a technique which can search and identify personal information in personal computer. Although people are familiar with personal computer and its usage, however they failed to be aware of the effect of unauthorized disclosure and manipulation of their personal information. They do not realise that unnecessary personal information in personal computer is in fact a liability. This project could help people to periodically check and clean personal information gathered in their personal computer respectively.

1.7 Project Organization

The organization of this project can be illustrated as figure 1.1 below.

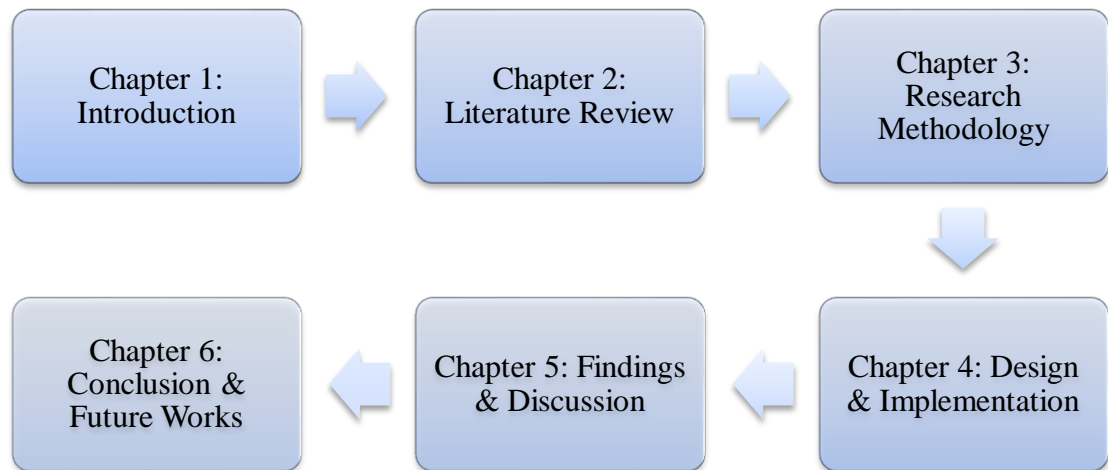


Figure 1.1: Project Organization

Chapter 1 which is the introduction has discussed about the background of the problem of this research. Introduction also includes the problem statement, purpose, scopes and objectives to be achieved. The next chapter is literature review. In literature review, the background information of the study is provided. Also detailed information of previous related works which reside on the same domain of this research is discussed. Chapter 3 is research methodology. Research methodology will discussed phases and process of developing this research. Chapter 4, the design and implementation will highlight the development of the technique. Chapter will discuss the obtain results for analysis and finally, chapter 6 will explain about future works recommendation and conclusion of this research.

REFERENCES

- (2004). "Chapter 9: Regular Expression." The Open Group Base Specifications 1003.1(6).
- (2008). "Sensitive Data Classification and Protection: Overcoming the Challenges to Classify and Protect Sensitive Data at Federal Government Agencies." SECURE IT White Paper.
- (2010). "The Case for Personal Information Empowerment: The rise of personal data store." Mydex CIC.
- (2010). "Standard for Environmental Assessment of Personal Computer Products, Including Notebook Personal Computers, Desktop Personal Computers, and Personal Computer Displays." IEEE Std 1680.1-2009: c1-33.
- (2012). "Computer Basics: Getting to Know Computers."
- Allan, R. A. (2001). "A History of Personal Computer: The People and the Technology."
- Anne Adams, M. A. S. (1995). "Users are not the Enemy."
- Authority, G. T. (2008). "Classification of Personal Information." GTA Information Security.
- Barnett White, T. (2004). "Consumer Disclosure and Disclosure Avoidance: A Motivational Framework." *Journal of Consumer Psychology* 14(1–2): 41-51.
- Beldad, A., et al. (2010). "How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust." *Computers in Human Behavior* 26(5): 857-869.
- Beldad, A., et al. (2011). "I trust not therefore it must be risky: Determinants of the perceived risks of disclosing personal data for e-government transactions." *Computers in Human Behavior* 27(6): 2233-2242.

- Boardman, R. and M. A. Sasse (2004). "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Vienna, Austria, ACM: 583-590.
- Erika McCallister, T. G., Karen Scarfone (April 2010). "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)." NIST National Institute of Standards and Technology.
- Garetz, M. (1999). "Evolutions of Microprocessor." 10: 209.
- Greenberg, S., et al. (1999). "PDAs and shared public displays: Making personal information public, and public information personal." *Personal Technologies* 3(1-2): 54-64.
- Intel (2011). "Protecting Sensitive Data on laptops is More Important Now Than Ever." White Paper, Intel Anti-Theft Technology.
- Johan P. van Braak, J. T., Martin Valcke (2005). "Explaining Different Types of Computers." *European Journal of Psychology Education*.
- Kibirige, G. W. (2012). "Awareness of Information Disclosure on Social Network Sites."
- Livingstone, S. (2008). "Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression." *New Media & Society* 10(3): 393-411.

The explosion in social networking sites such as MySpace, Facebook, Bebo and Friendster is widely regarded as an exciting opportunity, especially for youth. Yet the public response tends to be one of puzzled dismay regarding a generation that, supposedly, has many friends but little sense of privacy and a narcissistic fascination with self-display. This article explores teenagers' practices of social networking in order to uncover the subtle connections between online opportunity and risk. While younger teenagers relish the opportunities to recreate continuously a highly-decorated, stylistically-elaborate identity, older teenagers favour a plain aesthetic that foregrounds their links to others, thus expressing a notion of identity lived through authentic relationships. The article further contrasts teenagers' graded conception of 'friends' with the binary classification of social networking sites, this being one of several means by which online privacy is shaped and undermined by the affordances of these sites.

LLC, W. P. "Demographics and Psychographics."

Marina Sokolova, Y. J., David Schramm (2012). "Text Mining for Personal Health Information on Twitter." IEEE Second Conference on Healthcare Informatics, Imaging and System Biology.

Milberg, S. J., et al. (1995). "Values, personal information privacy, and regulatory approaches." *Commun. ACM* 38(12): 65-74.

Moon, Y. (2003). "Don't Blame the Computer: When Self-Disclosure Moderates the Self-Serving Bias." *Journal of Consumer Psychology* 13(1-2): 125-137.

Nagatomi, Y., et al. (2010). A Regular Expression Processor Embedded in Service-Friendly Router for Future Internet. *Parallel Processing Workshops (ICPPW)*, 2010 39th International Conference on.

Norberg, P. A., et al. (2007). "The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors." *Journal of Consumer Affairs* 41(1): 100-126

Nosko, A., et al. (2010). "All about me: Disclosure in online social networking profiles: The case of FACEBOOK." *Computers in Human Behavior* 26(3): 406-418.

Phelps, J., et al. (2000). "Privacy Concerns and Consumer Willingness to Provide Personal Information." *Journal of Public Policy & Marketing* 19(1): 27-41.

The authors examine potential relationships among categories of personal information, beliefs about direct marketing, situational characteristics, specific privacy concerns, and consumers' direct marketing shopping habits. Furthermore, the authors offer an assessment of the trade-offs consumers are willing to make when they exchange personal information for shopping benefits. The findings indicate that public policy and self-regulatory efforts to alleviate consumer privacy concerns should provide consumers with more control over the initial gathering and subsequent dissemination of personal information. Such efforts must also consider the type of information sought, because consumer concern and willingness to provide marketers with personal data vary dramatically by information type.

Society, I. C. (2010). "IEEE Standard for Environmental Assessment of Personal Computer Products, Including Notebook Personal Computers, Desktop Personal Computers and Personal Computer Display."

Sweeney, L. (2000). "Simple Demographics Often Identify People Uniquely." Data Privacy Working Paper 3.

Taraszow, T., et al. (2010). "Disclosure of personal and contact information by young people in social networking sites: An analysis using Facebook profiles as an example." *International Journal of Media & Cultural Politics* 6(1): 81-101.

In the context of the European Safer Internet project EU Kids Online, the aim of this article is to address how young people deal with privacy issues in social networking sites, using Facebook as an example. The study on which it is based examined the type of personal and contact information young people disclose through their profiles. In addition, it assessed gender differences in the disclosure of personal and contact information. A hundred and thirty-one Facebook member profiles were observed, selected to fit the European Commission's youth age range of 13-30. Results suggested that most people regardless of gender enter full name, facial pictures, hometown and e-mail addresses in their profiles. However, males are more likely than females to disclose mobile phone number, home address and instant messaging (IM) screen names. Consistent with the past literature, youth, especially between the ages of 18 and 22, seem unaware of the potential dangers they are facing when entering real personal and contact information in their profiles while accepting friendship requests from strangers. Recommendations for future research include investigating the levels of awareness young people have when disclosing information about themselves that can potentially harm them in more ways than one.

Turn, R. (2008). "Classification of Personal Information for Privacy Protection Purposes."

Van Eyll, C., et al. (1995). High performance regional wall synchrony analysis in severe systolic dysfunction: a new program based on reverse Polish notation. *Computers in Cardiology* 1995.

W. Burks, D. W. W. a. J. B. W. (2010). "An Analysis of a Logical Machine Using Parenthesis-Free Notation." *Mathematical Tables and Other Aids to Computation* Vol 8: 53-57.

Welch, G. W. a. M. (1985). "A Microcomputing Timeline." 10: 198.

Wilkes, M. V. (2000). "The Past and Future Development of Personal Computers." Olivetti Research, Cambridge.

Yoo, I. J. a. H. S. (2009). "Personal Information Classification for Privacy Negotiation." 2009 Fourth International Conference on COmputer Sciences and Convergence Information Technology.

Yu-Mei, N., et al. (2009). The optimization of RPN criticality analysis method in FMECA. Apperceiving Computing and Intelligence Analysis, 2009. ICACIA 2009. International Conference on.

New South Wales Information and Privacy Commission

http://www.ipc.nsw.gov.au/privacy/privacy_forgovernment/govt_privacy/privacy_faqpersonalinfo.html

Wyndowe v. Rousseau, 2008 FCA 39 (CanLII), <<http://canlii.ca/t/1vnhw>> retrieved on 2013-08-29

Privacy Commissioner of Canada

http://www.priv.gc.ca/leg_c/interpretations_02_e.asp#_ftn43

Privacy Commissioner's Report of Findings – Law School Admission Council Investigation – May 29, 2008 - http://www.privcom.gc.ca/cf-dc/2008/cf-dc_rep_080529_e.asp

PIPEDA Case Summary #1 - Video surveillance activities in a public place - http://www.privcom.gc.ca/cf-dc/2001/cf-dc_010615_e.asp.

PIPEDA Case Summary #351 - Use of personal information collected by Global Positioning System considered – http://www.privcom.gc.ca/cf-dc/2006/351_20061109_e.asp.

PIPEDA Case Summary #319 - ISP's anti-spam measures questioned - http://www.privcom.gc.ca/cf-dc/2005/319_20051103_e.asp