Assessing Trust in E-Commerce Website based on Ranking of Trust Attributes

Muhammad Rushdi Rusli [1], Ab. Razak Che-Hussin [2], Halina Mohamed Dahlan [3]

Information Systems Department
Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia 81310 Skudai, Johor, Malaysia.
[1]mrushdi84@gmail.com, [2]abrazak@utm.my, [3]halina@utm.my

Abstract – Trust has become one of most importance criterion in order to establish the e-Commerce. Previous study has shown that, there are five most important trust attributes that should be placed in an e-Commerce website. However, consumers may not know about the trust attributes and incapable to search the attributes manually especially for the new consumers. Besides, there are no proper tools available that can help the consumers to search all five trust attributes in an e-Commerce website. In order to overcome the problem, tool that can search the trust attributes should be develop to help the customer to assess the trustworthiness level of an e-Commerce website. Since this study only focus on e-Commerce websites in United Kingdom (UK), United States (US) and Malaysia, data from the websites have been collected to create patterns using regular expressions. Observation has shown that most of trust attributes are located in "Homepage" and "Contact us" pages. The proposed patterns then are applied into the prototype and the practicability of the prototype has been tested. ROC has been used to assign weight to each attributes depending on the rank of the importance of the trust attributes to help the customer to assess and making a decision about the trustworthiness of an e-Commerce website. Precision, recall and f-measure have been used to measure the performance of the prototype and the results show that the system performs almost same result by using manual search.

Keywords – Information Extraction, Regular Expression, e-Commerce, Trust

1 Introduction

Internet is creating a universal technology platform for e-Commerce and for driving important business processes inside organizations [1]. E-Commerce has numerously been employed as another virtual channel which is different from traditional brick-and-mortar channel in order to create business opportunity [2]. Therefore, an important measure for progressive e-Commerce is to help customs and companies fundamentally understand what e-Commerce would mean to them and how they should prepare for it [3]. Recent statistics from the Census Bureau of the Department of Commerce [4] indicate that total e-Commerce sales for 2008 were estimated around USD 133.6 billion, which is a 4.6 percent increase over the sales for 2007. However, an e-Commerce transaction is different from traditional commerce whereby consumers could not touch the target products, feel the shopping environment, contact salespersons for face-to-face inquiries, and take immediate possession upon purchasing [5]. To gain the customer's trust is an important task for the seller in order to secure an e-Commerce transaction. As a result, many people have focused upon trust issues in e-Commerce including researchers. Some of them have conducted studies on factors that influence consumer trust of e-Commerce websites. These factors are defined by the researchers with various definitions like trust influencers [6], antecedents of trust [7] and trust attributes [8] and each of them would have their own point of view.

Kim et al. [9] claimed that there are four categories of antecedents of trust which are cognition-based, affect-based, experience-based and personality-oriented. Similarly, McKnight et, al. [10] also developed a model of e-Commerce trust and categorized them as disposition to trust (consisting of faith in humanity and trusting stance), institution-based trust (consisting of situational normality and structural assurance), trusting beliefs, trusting intentions, personal innovativeness, web experience and perceived site quality. In contrast with both studies discussed previously, Che-Hussin et al. [8] have defined the trust factor with definition of trust attributes which are more firmer than from other studies. The trust attributes will be mentioned in the next section. However, it is highly probable that a new consumer is incapable of gathering sufficient information and assessing the quality and reputation based only on the web content provided by the vendor itself [5]. The focus of this study is how the trust attributes will be represented in regular expression patterns and develop a prototype that apply the proposed patterns to search the trust attributes.

This paper is organized as follows. In section 2, we describe the five trust attributes that exist in e-Commerce websites. Followed with the location of the trust attributes in e-Commerce. Section 3 describes the trust attributes patters and how to write patterns using regular expression. The patterns of data are gathered from e-Commerce websites within the UK, US and Malaysia. Once the patterns are created, search algorithm is proposed and applied in the system to search the five trust attributes. Furthermore, the prototype is tested on the e-Commerce websites from UK, US and Malaysia and comes out with the comparison between a manual search

and system search for the attributes. Based on the results, the patterns can be used to search the trust attributes which will somehow help the customer to gain trust towards e-Commerce website. The paper ends with the conclusion and suggestion for future work.

2 Existence of Trust Attributes in E-Commerce Website

According to the online questionnaires administered by Che-Hussin et al. [8], there are five trust attributes that should be placed on the first page of e-Commerce website to gain the trust of consumers. Table 1 shows the five trust attributes.

Table 1: Top five trust attributes adapted from Che-Hussin et al. [8]

Rank	Trust attributes
1.	Company telephone number (CTN)
2.	Company email address (CEA)
3.	Privacy Policy (PP)
4.	Company address (CA)
5.	Third party for secure transaction (e.g. VeriSign) (TPST)

Based on the attributes in Table 1, company telephone number and company email address are the most important attributes that will influence trust of customers towards an e-Commerce website. The attributes are important for customers to contact the vendor if they have any problem with the product that they purchased. These trust attributes are followed by the privacy policy which states the policy of the vendor regarding their products or services, the company address and third party for secure transaction. The last trust attribute is to guarantee that their personal information is safe and secure.

There are only five trust attributes that have been considered in this study since these attributes are the most important information for the consumer [8]. As we know, there many pages in an e-Commerce website and the trust attributes only exist in several of them. The search time for the system to find the trust attributes can be reduced by searching in the appropriate page only. The next section will explain how the locations are identified and what the common pages of the merchants who placed the trust attributes.

2.1 Location of Five Trust Attributes

The internet holds many e-Commerce. Since there is no standard way to present the trust attributes in the e-Commerce websites, there are possibilities that these trust attributes are located in the different places. A survey was conducted to find the common place of trust attributes in e-Commerce website. The observation has been done on ten UK, US and Malaysian e-Commerce websites.

Table 2: UK's e-Commerce websites

URL	Trust Attributes					PP Keyword Used
	CTN	CEA	PP	CA	TPST	
http://www.onedropperfumes.com/	CU	CU	NA	CU	NA	NA
http://mumnbaby.com	CU	CU	NA	CU	NA	NA
http://www.alicewonders.com/	CU	CU	HP/ CU	CU	NA	Privacy Policy
http://www.computermalaysia.com	NA	HP/CU	NA	CU	NA	NA
http://www.cardia.com.my/	CU	HP/CU	NA	CU	NA	NA
http://fashionstore.com.my/	CU	CU	HP/CU	CU	NA	Privacy Notice
http://www.goeskincare.com/	CU	CU	NA	CU	NA	NA
http://www.rcplanet.com.my/	CU	CU	HP/CU	CU	NA	Privacy Notice
http://www.hobbysportz.com	CU	NA	HP/CU	CU	NA	Privacy Notice
http://www.beautyimpress.com	CU	CU	HP/CU	CU	NA	NA

In Table 2, Table 3 and Table 4, CU stands for Contact Us, HP stands for Homepage and NA stands for Not Available. According to these tables, most e-Commerce websites from UK, US and Malaysia place their trust attributes in HP and CU. However, most of them also do not have any third party for secure transaction especially in the US and Malaysia. These attributes and privacy policy usually exist in HP and CU. Besides, e-Commerce websites always put their privacy policy under "Privacy Policy", "Privacy & Security", "Privacy and Security", "Privacy Notice", "Privacy", or "Policies". This section, contains details of what personal information is collected, how the personal information may be used, the persons to whom the personal information may be disclosed, the security measures taken to protect the personal information, and so on. Sometimes, personal information of the customer is needed in order to conduct an e-Commerce transaction.

Table 3: US's e-Commerce websites

URL	Trust Attributes					PP Keyword Used
	CTN	CEA	PP	CA	TPST	
http://www.buy-jeans.net/	HP/ CU	HP/CU	NA	CU	NA	NA
http://www.customwaxnseals.co.uk	HP/CU	HP/CU	NA	HP/CU	NA	NA
http://www.distinctlybritish.com	HP/CU	CU	NA	HP/CU	HP/CU	NA
http://www.elc.co.uk/	HP	NA	HP	HP	NA	Privacy Policy
http://www.epicheroes.com	CU	CU	HP/ CU	CU	HP	Privacy Notice
http://www.framarhealth.com	HP/ CU	CU	NA	CU	NA	NA
http://www.majestic.co.uk/	CU	CU	HP/CU	CU	NA	Privacy Policy
http://www.jewellerynow.co.uk/	CU	CU	HP/ CU	CU	NA	Privacy
http://www.simplysalmon.co.uk	HP/ CU	CU	NA	CU	NA	NA
http://www.shoe-shop.com/	CU	CU	HP/ CU	CU	NA	Privacy and Security

Table 4: Malaysia's e-Commerce websites

URL	Trust Attributes					PP Keyword Used
	CTN	CEA	PP	CA	TPST	
http://www.americanmedical-id.com/	HP/ CU	NA	HP/ CU	CU	HP/ CU	Privacy Policy
http://www.scholarships.com	NA	NA	HP	CU	HP	Privacy Policy
http://www.pageonce.com/	CU	CU	HP/ CU	CU	HP/ CU	Privacy & Security
http://www.diapers.com/	HP/ CU	HP/ CU	HP/ CU	HP/ CU	HP/ CU	Security & Privacy Policy
http://www.babyearth.com/	CU	CU	HP/ CU	NA	HP/ CU	Privacy Policy
http://www.unbeatablesale.com/	HP /CU	CU	HP/ CU	CU	NA	Policies
http://www.nurserydepot.com	CU	CU	HP/ CU	CU	NA	Privacy Policy
http://www.iseeme.com/	CU	CU	HP/ CU	CU	HP/ CU	Privacy Policy
http://www.ignatius.com	HP	HP	HP	HP	NA	Privacy Policy
http://www.bookcloseouts.com/	CU	CU	HP/ CU	CU	HP	Privacy Policy

Since this information is private and confidential, the company should have a policy to make sure the customers' information is not misused by any organization.

Besides that, there are also possibilities that the information can be stolen by any irresponsible party. Merchants use third party secure transaction to prevent such thing from happening. Nowadays, there are many third parties that offer secure transactions services which will be mentioned in the next section. Normally, the merchants put the service logo in the homepage section.

Once trust attributes and all locations of the attributes are identified, another issue that arises is how to make the system recognize the attributes in order to extract it. Since these trust attributes are usually located in unstructured text, information extraction technique is used to extract them by using patterns to recognize the desired data. This study used regular expression to create patterns of the trust attributes which are explained in the next sections.

3 Converting Trust Attributes to Regular Expression Patterns

Even though there are five trust attributes, only four patterns are created for CTN, CEA, CA and TPST. This study only searched links for PP attributes to identify whether an e-Commerce website contained this attribute. The next section will explain how the patterns are created for each trust attribute.

Before patterns are created using regular expression, the pattern of the desired data should be recognized first. In order to recognize the pattern, a survey was conducted on several e-Commerce website. Since this study only focus on e-Commerce websites from United States, United Kingdom and Malaysia, the samples of data are taken from those countries only. The survey was conducted on 40 e-Commerce websites from United States, 40 websites from United Kingdom and 40 websites from Malaysia. Table 5 shows five e-Commerce websites out of 40 patterns of CA and CTN from these countries since these attributes are different amongst the three countries mentioned compared with CEA, PP and TPST attributes.

Regular expression also known as regex is a common way to describe a pattern of characters whether match or fail to match, sequences of characters in text [11]. It allows the computer users and system developer to find the desired text in a string and sometimes it is used to replace certain text with other text. Regular expression is also used to validate the inputs that have been inserted by the user are suitable for storage. The main component in regular expression is to find the match pattern in a string.

Since most of the e-Commerce websites are in unstructured format, these trust attributes are also written and placed in this format. One of the solutions to search these attributes in unstructured text is by using regular expression. Based on the patterns of data in Table 5, trust attributes patterns are derived into regular expressions patterns.

Table 5: Example Patterns of CA and CTN among UK, US and Malaysia

Country	No	URL	Trust Attributes	
			CA	CTN
UK	1.	http://www.distinctlybritish.com	Distinctly British Limited 20-22 Bedford Row London WC1R 4JS	0870 428 5488
	2.	http://www.customwaxnseals.co.uk	The Design Station Ltd., 9 Turnstone Drive, Featherstone, Wolverhampton, WV10 7TA, England	+44 (0) 7944 535244
	3.	http://www.framarhealth.com	595 Lisburn Road, Belfast, Co. Antrim,N Ireland BT9 7GS	+44 (0) 28 9068 1015
	4.	http://www.jewellerynow.co.uk	Jewellery Now, Devonshire House, Manor Way, Borehamwood Herts WD6 1QQ	08700 466420
	5.	http://www.caldertrade.co.uk/	Units 11 & 12, Halifax Industrial Centre, Pellon Lane, Halifax, West Yorkshire, HX1 5RW.	(01422) 330008
US	6.	http://www.pageonce.com	El Camino Real Palo Alto, CA 94306 USA	650.560.6500
	7.	http://www.diapers.com/	Diapers.com, P.O. Box 1564, Montclair, NJ 07042	1-800-342-7377
	8.	http://www.minitots.com	25 Robert Pitt Drive #106, Monsey NY 10952	866-307-8687
	9	http://www.magazines.com	Magazines.com LLC P.O. Box 682108 Franklin, TN 37068	1-800-624-2946
	10.	http://www.liquidation.com/	1920 L Street, NW, 6th Floor, Washington, DC 20036	(800) 310-4604

Malaysia	11.	http://www.rcplanet.com.my/	RC Planet, Taman Permainan, Jalan SS19/1, 47500 Subang Jaya, Selangor, Malaysia.	+60 3 5636 5163
	12.	http://www.hobbysportz.com	No 3A-G, Jalan Cempaka SD12/1, Bandar Sri Damansara 52200 Kuala Lumpur Malaysia	012-2361337
	13.	http://www.rcsmart.com.my/	Lot 1 - 47A , 1st Floor, Amcorp Mall, 18, Jalan Persiaran Barat, 46050 Petaling Jaya, Selangor, Malaysia	+603 79600028
	14.	http://www.goeskincare.com/	B2-02, PJ Industrial Park, Jalan Kemajuan, 46200 Petaling Jaya, Selangor	+603-7954 7877
	15.	http://www.onedropperfumes.com/	Suite 18-4, UBSA Tower, Persiaran Damai, Section 11, Shah Alam, 40100, Selangor, Malaysia	+603-55100441

4.1 Company Telephone Number (CTN)

In Table 6, number 1 and 2 regular expressions are used to extract company telephone numbers from United Kingdom, number 3 for United States, number 4 for United States and United Kingdom while number 5 and 6 for Malaysia.

Table 6: Regular expression patterns to CTN

Country	Regular Expression
UK	**Pattern (a):** \(?\d{5}\)?[-\s.]\(?\d{6}\)? Example: i. (01422) 330008 ii. 08700 466420 **Pattern (b):** \+?\d{2}[-\s.]\(?\d{1}\)?[-\s.]?\d{0,3}[-\s.]\d{3,4}[-\s.]\d{3,6} Example: i. +44 (0) 28 9068 1015 ii. 44 (0)20 8960 6900 iii. +44 (0) 1753 833334
US	\(?\d{3,4}\)?[-\s.]\d{3}[-\s.]\d{4} Example: (800) 564-5740
UK & US	\d{1}?[-\s.]?\+?\(?\d{2,3}\)?[-\s.]\d{3,4}[-\s.]\d{3,4}[-\s.]?\d{0,4} Example: i. 1-888-237-8289, 650.560.6500 ii. +44 844 844 0809 iii. +44 1829 771 886
Malaysia	**Pattern (a):** \+?\d{3,4}[-\s.]\d{7} Example: i. +6016-6242492 ii. 016-4317007 **Pattern (b):** \+?\(?\d{2,4}\)?[-\s.]\d{4}[-\s.]?\d{3,4} Example: i. +603-8922 1513 ii. (603) 2938 3818 iii. 03-55694549

3.1.1. CTN patterns for UK

Pattern (a) in Table 6 is used to find UK company telephone numbers. In regular expression, backslash character '\' is used to match a special character. For example, in order to match the telephone numbers which starts with five digits and followed by whitespace and 6 digits. Some of them write the first five digits between left and right parenthesis. Since character '(' and ')' are categorized as special characters, the character '\' should be placed first. The character '?' is to show that the character before it can exist or not because to match both strings that have character '(' and ')' or not. The character '\d{5}' is used to match digit characters as much as five characters. The pattern '\(?\d{5}\)?' can be used to match the first string of telephone numbers such as (01422) or 08700. The character '[-\s.]' is used to match character '-', '.' or whitespace. And the pattern '(?\d{6}\)?' is the same as the previous which is used to match digit characters as much as six characters.

Pattern (b) in Table 6 is used to search for UK company telephone numbers. In order to match character '+44' or '44', pattern \+?d{2} is used. Followed with '[-\s.]' pattern to match character '-', '.' or whitespace. As in example 2, there are also data that is without whitespace in '(0)20' which is different with '(0) 28', the pattern '[-\s.]' should include the '?' character. The pattern 'd{0,3}' is to match digit number which is not more than 3 digit and can be 0 digit. This is used for the string in example 3 where the arrangement of the digit is different from the first two examples. According to the example 1 and 2, the arrangement of the digit is 2 digits, 4 digits and followed by 4 digits. However, in the last example, the arrangement of the number is 4 digits followed by 6 digits.

3.1.2. CTN patterns for US

The pattern in the US category in Table 6 is used to find the company telephone numbers from US. Pattern '\(?\d{3,4}\)?' is to match the first three digits '(800)' and '[-\s.]\d{3}' is to match the whitespace and three digits in the middle of the string. Pattern '[-\s.]\d{4}' is to match the hyphen character (-) and the last four digits.

3.1.3. CTN patterns for both US and UK

The pattern in the US and UK categories in Table 6 is created to match telephone numbers from UK and US since the pattern of the arrangement of number is similar. The pattern is easier to create when the pattern of the string is similar with the other string which is desired to match. Pattern '\d{1}?[-\s.]?\+?\(?\d{2,3}\)?' is to match the string '1-', '650.', or '+44'. The rest of the pattern is to match the number which follows with three or four digits such as '888-237-8289', '560.6500', '844 844 0809' or '1829 771 886'. According to the example, the vendors place their telephone numbers in various formats. Some of them use character hyphen '-', periods '.' and whitespace (' ').

3.1.4. CTN patterns for Malaysia

There are only two types of telephone numbers in Malaysia which are mobile phone and fixed line. Pattern (a) for Malaysia category in Table 6 is used to match mobile phone numbers where several e-Commerce websites in Malaysia also includes number '60' as the country code. The number will follow with two digits, most of them are service provider codes, and followed with seven digits.

Meanwhile, pattern (b) is used to match fixed line telephone numbers in Malaysia where the numbers are placed in various ways such as inclusive of country code and the plus '+' character, and some of them put the country code in parenthesis '()'.

4.2 Company Email Address (CEA)

This trust attribute follows a standard pattern for the e-Commerce companies from UK, US and Malaysia. Table 7 shows the regular expression to extract email address.

Table 7: Regular expression patterns to extract company email address (applies for all)

Regular Expression
/^[A-z0-9\-_]+(\.[A-z0-9\-_]+)*@(([A-z0-9]+\-?[A-z0-9]+)+\.)+[A-z]{2,6}$/

This pattern is used to match the email pattern. The first pattern '/^[A-z0-9\-_]+(\.[A-z0-9\-_]+)*@' is to find local name which may only contains letters, digits, hyphen '-', underscore '_' and periods '.'. The local name may not begin and/or end with a period and not contain two or more subsequent periods '..'. Pattern '(([A-z0-9]+\-?[A-z0-9]+)+\.)' is use to match sub domain and domain of the email where the domain name and sub-domain name may not begin and/or end with a hyphen, must be between 2 to 63 characters long, and may only contain letters, digits and hyphen. Besides, top-level domains may only contain letters and must be between 2 to 6 characters long and the sub-domain names, the domain name and the top-level domain names are separated by single periods '.'.

4.3 Privacy Policy (PP)

According to Table 2, Table 3 and Table 4, e-Commerce websites from UK, US and Malaysia put their privacy policies under "Privacy Policy", "Privacy & Security", "Privacy and Security", "Privacy Notice", "Privacy", or "Policies" section. The method used by authors to identify whether e-Commerce website from UK, US and Malaysia include a privacy policy was to find the title of each link in the homepage that contains the words "privacy" or "policies". Since this method was only to find these words for the title link, deriving regular expressions pattern for this attribute is not necessary.

4.4 Company Address (CA)

Regular expression in Table 6 can only be used to extract company address from US and UK.

Table 8: Regular expression patterns to extract company address for United Kingdom and United States

Country	Regular Expression
US	/[A-Z]{2}[\s][ ]{0,},?[\s]{0,}\d{5}/
UK	/[A-Z]{1,2}[0-9]{1,2}[\s][0-9]{1}[A-Z]{1,2}/

3.4.1. CA patterns for US

Based on the observation of US e-Commerce websites, most of them write their address with state code followed by with the zip code. The state code usually is in capital letters with two letters and the zip code in five digits. Pattern '[A-Z]{2}' is to match the first two letters of the state code. Pattern '[\s][
]{0,},?[\s]{0,}' for US category in Table 8 is to match the character that separates the state code and zip code which is usually a whitespace ' ', html code for new line '
' and comma ','. Pattern '\d{5}' is used to match the zip code.

3.4.2. CA patterns for UK

UK address format is more specific than US address format where the codes are divided into postcode area, postcode district, postcode sector and postcode unit. The first letter or pair of letters represents the postcode area. The following number, from 0 to 99, determines the postcode district within that area. It follows with the whitespace ' '. After that the first character after the space is a digit from 0 to 9 which determines the postcode sector. The final two letters is for the postcode unit. Table 8 in UK category shows the regular expression to extract address from UK website.

3.4.3. CA patterns for Malaysia

Since Malaysian addresses do not have any state code, to extract the address is a difficult task. To overcome this problem, this study stored information that includes city and the state from the country into the database. The system will search state name then search the city of the state in 60 characters before the state's name is found.

Table 9: Regular expression patterns to extract company address for Malaysia

Type	Regular Expression
State	/$state/i
City	/$city/i
Postcode	/\d{5}

Besides that, the system will also search for the postcode if there is no city for the state found in 60 characters before the state name is found. According to Table 5, a state like 'Kuala Lumpur' does not have a city in the address. In this case, postcode is used to determine whether the state's name is found in address or not. '$state' is for the state name and '$city' is for the city name.

4.5 Third Party for Secure Transaction (TPST)

Regular expression in Table 10 shows regular expression patterns to extract eight third party secure transaction information from an e-Commerce websites. This study only focuses on these eight third party since they are commonly used by the e-Commerce company. According to the table, three of them constitute more than one regular expression since the pattern of the data is different from the others.

In order to identify whether an e-Commerce website does have third party secure transaction, each pattern of the link for third party secure transaction should be identified first. To obtain the pattern, an observation was conducted on 40 e-Commerce websites from United States, 40 websites from United Kingdom and 40 websites from Malaysia that have this service. Some of the websites are shown in Table 5.

Table 10: Regular expression patterns to extract third party secure transaction information.

Name	Pattern
GEOTRUST	/\/\/smarticon.geotrust.com\/si\.js/i
VERISIGN	a. /https:\/\/seal.verisign.com+[a-zA-z0-9=?\/;_&.]*/i b. /https:\/\/servicecenter.verisign.com\/cgi-bin\/Xquery\.exe\?/i
SCAN ALERT	a. /https:\/\/www.scanalert.com+[a-zA-z0-9=?\/;_&.]*/i b. /https:\/\/www.mcafeesecure.com\/RatingVerify\?ref=+".$urlsecure."/i
GLOBALSIGN	/https:\/\/secure.globalsign.net\/en\/find\/sealct.cfm\?id=\d+/i
TRUSTWAVE	/https:\/\/sealserver.trustkeeper.net\/compliance\/cert.php\?code=\w+/i
WebSafeShield	/http:\/\/seals.websafeshield.com\/\w+\/websafeshield\.js/i
Entrust	/https:\/\/seal.entrust.net\/seal\.js\?domain=".$urlsecure."/i
Thawte	a. /https:\/\/siteseal.thawte.com\/cgi\/server\/thawte_seal_generator\.exe/i b. /https:\/\/siteseal.thawte.com\/cgi\/server\/certdetails\.exe\?code=\w+/i c. /https:\/\/www.thawte.com\/cgi\/server\/certdetails\.exe\?referer=/i

According to the observation, some of the links for the secure transaction are not fixed to only one pattern but have more than one such as VERISIGN, sCAN ALERT, and Thawte. Therefore, the regular expressions patterns for these parties constitute more than one regular expression pattern. These patterns are created based on their link code that is used by the third party of secure transaction and utilized by their clients.

4.6 Search Algorithm

Previous sections discussed trust attributes that affect consumer's trust toward e-Commerce websites in order to establish the transaction. Homepage and Contact Us page have been identified as the location that the merchant used to placed their trust attributes in their website. Sample data are taken from several e-Commerce websites to create patterns for each trust attribute. Finally, all the information is used to generate a search algorithm to search the attributes.

The steps to search five trust attributes in an e-Commerce Website that has been proposed in the prototype are as figure 1. Figure 2 the flow of the search algorithm.

1. Convert the inserted URL into source code.
2. If success
   - 2.1. Check link's title contains word "privacy" or "policy"
   - 2.2. Check whether the homepage contains CTN, CEA, CA and TPST.
   - 2.3. Found CTN, CEA, CA and TPST?
     - 2.4.1. True
       Go to end
     - 2.4.2. False
       - 2.4.2.1. Search link's title contains word "contact" in the homepage
       - 2.4.2.2. Repeat step 2.3
3. Else
   Go to end

Fig. 1: Steps to search five trust attributes

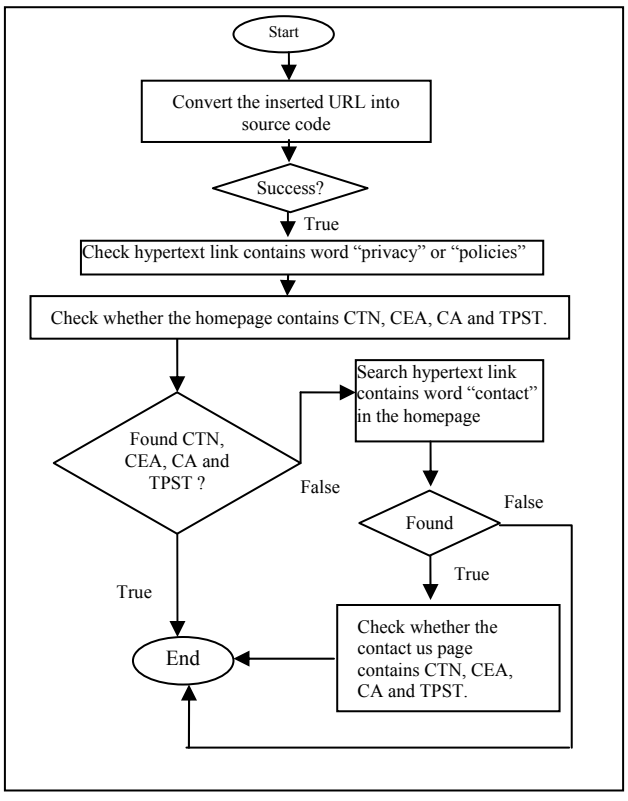

Fig. 2: Search Algorithm Flow

According to Fig. 2, there are several steps to search five trust attributes in an e-Commerce Website. Before the processes begin, the user has to insert the URL of the desired e-Commerce Website in order to assess the

trustworthiness of the Website. The system will convert the inserted URL into source code. If the process is a success, the system will search all the links and the name of the link in the page that contains the words "privacy" or "policies". If the name of the link contains either of these two words, it means that the Website contain privacy policy trust attributes. For the four trust attributes which are company telephone number, company email address, company address and third party secure transaction, the system will search the attributes in the homepage that has been inserted with the URL page first. If it cannot find the attributes in the page, the system will search the link's title that contains the word "contact" in that page. According to the Table 2, Table 3 and Table 4, most e-Commerce Website placed their information such as email, address, telephone number and third party secure transaction in 'Homepage' and 'Contact Us' page.

4.7 Trustworthiness Assessment

Every e-Commerce website has its level of trustworthiness for consumers to assess how deep they can trust an e-Commerce website. There should be a mechanism to help the customer to make a decision whether to trust or not trust an e-Commerce website. This study has set a different weight for each every trust attribute depending on the rank of the importance of the trust attributes.

Since this study focus on five trust attributes, multi attributes decision analysis is used to assign weight to every attributes. There are several methods to assign weight to attributes such as Equal Weighting method, Ratio Weights method and Rank-order method. Each method has their advantages which are depend to the information that the decision maker has about the attributes.

Equal Weighting (EW) method which suggested by Dawes and Corrigan [12] can be used when the rank information is inconsequential [13]. Each of attributes will be assigned with the same weight and best on average when there is no information about the weights [14]. This method preserves only categorical information from a decision maker's judgments– either an attribute matters or it does not [15].

Ratio weights methods such as Swing Weights and Trade-off weights preserve ratio scale properties of the decision maker's judgments [15].The weight of attributes are depend to the range which is the greater the range of outcomes for an attributes, the greater the weight for the attribute [16].

Rank-order methods rely only on ordinal information about attribute importance and provide a superior approach [15]. Some of the methods that include in this category are the Rank-Sum (RS) and Rank-Order-Centroid (ROC) methods. With these methods, ratios among weights are established by applying a standard transformation of ranks into ratios.

In order to assign weight to each trust attributes, ROC has been selected since the weight is assign based on the importance of rank of an attributes which is suitable to apply to this study. According to Jia et al. [15], when the rank ordering of attributes was known with certainty, the ROC method is the best of the approximate methods especially when the number of attributes is large. ROC method is selected in this study to calculate and assign the weight for each trust attributes to help the customer to assess and making a decision about the trustworthiness of an e-Commerce website. In general, for the i th most important attribute the centroid weight is:

$$\omega_i(ROC) = \frac{1}{n}\sum_{j=i}^{n}\frac{1}{j}, i = 1, \dots, n$$

This study has five trust attributes that follow in rank order are CTN, CEA, PP, CA and TPST. Below is the calculation to assign weight to each of trust attributes. Table 11 shows the weight of the trust attributes.

Number of trust attributes, n = 5,

W(1)=CTN;W(2)=CEA;W(3)=PP;W(4)=CA;W(5)=TPST

W(1) = (1+1/2+1/3+1/4+1/5)/5 = 137/300;
W(2) = (1/2+1/3+1/4+1/5)/5 = 77/300;
W(3) = (1/3+1/4+1/5)/5 = 47/300;
W(4) = (1/4+1/5)/5 = 27/300 ;
W(5) = (1/5)/5 = 12/300

Table 11: Weight of the Trust Attributes

Rank	Trust attributes	Weight (%)
1.	Company telephone number	45.67
2.	Company email address	25.67
3.	Privacy Policy	15.66
4.	Company address	9
5.	Third party for secure transaction (e.g. VeriSign)	4

The weight will be used to calculate the level of trustworthiness of an e-Commerce website. For example, an e-Commerce website that contains only company telephone number is better than the website that contains only company email address, company address and third party secure transaction since the weight of company telephone number is 45.67% which is much greater than the total weight of company email address, company address and third party secure transaction which add up to 38.67% only.

4.0 Comparison between Manual Search and System Search

After the patterns have been created using regular expression, these patterns are finally applied in a web based system to search the five trust attributes. The comparison between manual search and system search is shown in Table 12 which is only shows 10 out of 40 each of e-Commerce websites from UK, US and Malaysia.

According to the table, the system is capable of finding the five trust attributes of almost all websites that have been tested. Some of trust attributes failed to be retrieved

because of six factors, such as no mention of state code in the website, spelling error of the state name, address that is written different from with the pattern that has been created, using image for the contact us link, put email information in image format and using another third party for secure transaction that is not included in this study. Figure 3 shows the result by using system search that has been conducted on 120 e-Commerce websites which each 40 websites from UK, US and Malaysia.

Fig. 3: System Search Result

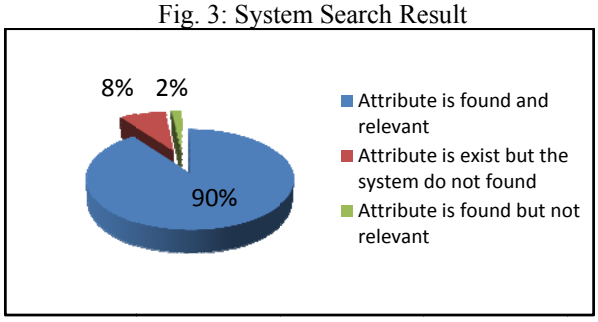

Fig 3 shows that 90% which is 377 from 419 of trust attributes are successfully been found by using the system. Only 10% where 8% (35) and 2% (7) from 419 of them are failed to be searched which happen for several reasons that are mentioned earlier.

4.1 Evaluation Method

This study evaluates the performance of the system search by using precision, recall and f-measure. This step is important to measure how precise and how complete the classification is on the positive class [17]. Table 13 shows the confusion matrix that contains information about actual and predicted results given by a classifier.

Table 13: Confusion matrix of a classifier

	Classified positive	**Classified negative**
Actual positive	TP	FN
Actual negative	FP	TN

Where;
TP: the number of correct classifications of the positive examples (true positive)
FN: the number of incorrect classifications of positive examples (false negative)
FP: the number of incorrect classifications of negative examples (false positive)
TN: the number of correct classifications of negative examples (true negative)

Based on the Table 12, the precision (P) and the recall (R) of the positive class are defined as follows:

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FN}$$

According to the Wilson [**18**], the precision is defined as the fraction of retrieved documents that are relevant and recall is defined as the fraction of relevant documents that are retrieved. Besides that, these measurements are hard to compare classifiers since they are not functionally related. Even they are not related to each other, somehow high precision is achieved at the expense of recall and high recall is achieved at the expense of precision [**17**]. F-measure also can be used to compare different classifier which is the harmonic-mean of precision and recall. The harmonic mean of two numbers tends to be closer to the smaller of the two. Its mean both P and R must be high for the F-measure to be high.

$$F - measure = \frac{2PR}{P+R}$$

4.2 Evaluation of Result

Below is the calculation to get the value of precision, recall and f-measure.

TP = 377 FP=7 FN=35

$$P = \frac{TP}{TP + FP} \qquad P = \frac{377}{377 + 7} \qquad \text{P = 0.98}$$

$$R = \frac{TP}{TP + FN} \qquad R = \frac{377}{377 + 35} \qquad \text{R = 0.92}$$

$$F = \frac{2PR}{P + R} \qquad F = \frac{2 \times 0.98 \times 0.92}{0.98 + 0.92} \qquad \text{F = 0.95}$$

According to the calculation, the search tool by applying the proposed patterns gives best performance which precision is 0.98, recall is 0.92 and f-measure is 0.95. The results indicate that the proposed patterns can be used to search five trust attributes in an e-Commerce website from UK, US and Malaysia with the value of precision, recall and f-measure quite high and almost same result by using manual search.

5.0 Conclusion and Future Works

It can be concluded that, trust is an important factor in order to establish e-Commerce transactions. Previous researchers have identified the trust attributes that should be placed in an e-Commerce website which are company telephone number, company email address, privacy policy, company address and third party for secure transaction. These trust attributes are the most important things to gain customer confidence towards e-Commerce websites. The study also shows that most of the websites placed the attributes in Homepage and Contact Us pages. Information extraction technique is used since almost all the information in the websites is in unstructured text. The technique used patterns of desired data to extract them. Sample data were taken from each 25 e-Commerce website from UK, US and Malaysia to create patterns of

the trust attributes. This study used regular expressions to write the patterns of the trust attributes. Once the patterns are created, a system that can search the attributes can be developed according to the search algorithm that has been proposed. This study is important since the task of searching the trust attributes is somehow hard especially for the beginner computer user and tools to help them in searching the trust attributes should be developed.

The experiment that has been conducted shows that regular expression can be used to write patterns in order to search the trust attributes. The future work that will be carried out shall include integrate with the machine learning method that capable to search trust attributes in e-Commerce websites besides from UK, US and Malaysia.

Symbol	Description
√	Exist
×	Do not exist
y	Attribute is found
χ	Attribute is not found
TA	Trust Attribute

Table 12: Comparison result between manual search and system search.

URL	Manual Search						System Search					Attribute that failed to be searched & the reason
	Trust Attributes						Trust Attributes					
	CEA	CA	CTN	PP	TPST		CEA	CA	CTN	PP	TPST	
US												
http://www.americanmedical-id.com/	×	√	√	√	√		×	χ	√	√	√	TA - CA Reason - No state code in address
http://www.scholarships.com	×	√	×	√	√		×	χ	×	√	√	TA - CA Reason - No state code in address
http://www.pageonce.com/	√	√	√	√	√		√	√	√	√	√	None
http://www.diapers.com/	√	√	√	√	√		√	√	√	√	√	None
http://www.babyearth.com/	√	×	√	√	√		√	×	√	√	√	None
http://www.unbeatablesale.com/	√	√	√	√	×		√	√	√	√	×	None
http://www.nurserydepot.com	√	√	√	√	×		√	√	√	√	×	None
http://www.iseeme.com/	√	√	√	√	√		√	√	√	√	√	None
http://www.ignatius.com	√	√	√	√	×		√	√	√	√	×	None
http://www.bookcloseouts.com/	√	√	√	√	√		√	χ	√	√	√	TA - CA Reason - Address format is different with pattern
UK												
http://www.buy-jeans.net/	√	√	√	×	×		√	√	√	×	×	None
http://www.customwaxnseals.co.uk	√	√	√	×	×		√	√	√	×	×	None
http://www.distinctlybritish.com	√	√	√	×	√		√	√	√	×	√	None
http://www.elc.co.uk/	×	√	√	√	×		×	√	√	√	×	None
http://www.epicheroes.com	√	√	√	√	√		√	√	√	√	χ	TA - TPST Reason - Use Comodo third party secure transaction
http://www.framarhealth.com	√	√	√	×	×		√	√	√	×	×	None
http://www.majestic.co.uk/	√	√	√	√	×		√	√	√	√	×	None
http://www.jewellerynow.co.uk/	√	√	√	√	×		√	√	√	√	×	None
http://www.simplysalmon.co.uk	√	√	√	×	×		√	√	√	×	×	None
http://www.shoe-shop.com/	√	√	√	√	×		√	√	√	√	×	None
Malaysia												
http://www.onedropperfumes.com/	√	√	√	×	×		√	√	√	×	×	None
http://mumnbaby.com/	√	√	√	×	×		√	√	χ	×	×	TA - CTN Reason - Use image instead of text for contact us page
http://www.alicewonders.com/	√	√	√	√	×		√	χ	√	√	×	TA - CA Reason - No state's name
http://www.computermalaysia.com	√	√	×	×	×		√	χ	×	×	×	TA - CA Reason - Use image instead of text for contact us page
http://www.cardia.com.my/	√	√	√	×	×		√	√	√	×	×	None
http://fashionstore.com.my/	√	√	√	√	×		√	χ	√	√	×	TA - CA Reason - Spelling error for state's name
http://www.goeskincare.com/	√	√	√	×	×		√	√	√	×	×	None
http://www.rcplanet.com.my/	√	√	√	√	×		√	√	√	√	×	None
http://www.hobbysportz.com	×	√	√	√	×		×	√	√	√	×	None
http://www.beautyimpress.com	√	√	√	×	×		χ	√	√	×	×	TA - CEA Reason - Email in image format

6.0 References

1. A. Isazadeh, F. Z. Milani and A. F. Moghaddan, *Automated update of e-commerce systems using dss*, WSEAS Transactions on Systems, 2004, pp. 2378-2382.
2. H.-J. Ho and R. Yang, *A comparison of secure mechanisms for mobile commerce*, Proceedings of the 7th WSEAS International Conference on Mathematics & Computers in Business & Economics, 2006, pp. 24-28.
3. H. Liu, Q. Peng and J. Shen, *A reference model for e-commerce*, WSEAS International Conference on Electronics, Control & Signal Processing, 2002.
4. The Census Bureau of the Department of Commerce, "Quarterly retail e-commerce sales," The Census Bureau of the Department of Commerce, 2009.
5. L.-T. Huang, K.-L. Yin and C.-K. Farn, "The influence of consumers' consciousness of institutional factors on initial trust building," *Proceedings of the 5th WSEAS International Conference on Telecommunications and Informatics*, Instambul, Turkey, 2006, pp. 360-365.
6. K. Jones, L. N. K. Leonard and C. K. Riemenschneider, *Trust influencers on the web*, Journal of Organizational Computing and Electronic Commerce **19** (2009), no. 3, pp. 196-213.
7. H. Gill, K. Boies, J. E. Finegan and J. McNally, *Antecedents of trust: Establishing a boundary condition for the relation between propensity to trust and intention to trust*, Journal of Business and Psychology, **19** (2005), no. 3, pp. 287-302.
8. A. R. Che-Hussin, L. Macaulay and K. Keeling, *The importance ranking of trust attributes in e-commerce website*, 11th Pacific-Asia Conference on Information Systems (2007).
9. D. J. Kim, D. L. Ferrin and H. R. Rao, *A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents*, Decision Support Systems **44** (2008), pp. 544-564.
10. D. H. McKnight, V. Choudhury and C. Kacmar, *Developing and validating trust measures for e-commerce: An integrative typology*, Information Systems Research **13** (2002), no. 3, pp. 334-359.
11. A. Watt, *Beginning regular expression*, Wiley Publisher, 2005.
12. R. M. Dawes and B. Corrigan, *Linear models in decision making*, Psychological Bulletin **81** (1974), pp. 95-106.
13. F. H. Barron and B. E. Barrett, *Decision quality using ranked attribute weights*, Management Science **42** (1996), no. 11, pp. 1515-1523.
14. F. H. Barron and B. E. Barrett, *The efficacy of smarter — simple multi-attribute rating technique extended to ranking* Acta Psychologica **93** (1996), no. 1-3, pp. 23-36.
15. J. Jia, G. W. Fischer and J. S. Dyer, *Attribute weighting methods and decision quality in the presence of response error: A simulation study*, Journal of Behavioral Decision Making **11** (1998), no. 2, pp. 85-105.
16. G. W. Fischer, *Range sensitivity of attribute weights in multiattribute value models* Organizational Behavior and Human Decision Processes **62** (1995), no. 3, pp. 252-266.
17. B. Liu, *Web data mining - exploring hyperlink, contents and usage data*, Springer, 2007.
18. M. L. Wilson, "Information retrieval (precision and recall)," Swansea University (Computer Science), 2009.