CONTENT TRUSTED SCALE: AN ALGORITHM TO EVALUATE THE TRUSTWORTHINESS OF A WEBSITE

WADEE ALHALABI

Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia
Department of Computer Science, Effat University, Jeddah, Saudi Arabia

Abstract- According to the minimum-security concepts of information security, the service provider of a website must ensure confidentiality, integrity, availability, non-repudiation, and access control. These concepts provide risk free access, but they cannot ensure truthfulness of the contents. Trust plays an important role in our life, whether it is with our colleagues, friends, or family. It is the same case when we use the Internet as an information system resource. We need to trust the information and avoid any spurious or incomplete resources. The Internet is used as an information system resource where billions of websites exist and exchange information. Sometimes we get confused regarding which information to accept and which to reject. Every site has a different style of presenting data, and when we surf the Internet, we find that the same topic is being discussed by number of sites using different ways to present data. Often, the data is contradictory. Therefore, we must have a novel algorithm to eliminate low quality and untrusted resources in order to provide quality contents for the readers. Many techniques have been presented to overcome trust issues but most of them only focus on authentication and reputation of the websites than the offered content. Consequently, in our paper, we expanded the work in order to build an algorithm that can be used online to gage the trust score of the content provided on a particular site and to minimize or quantify the errors and faulty information given to users. Using recommender systems and semantic web approach, the proposed algorithm will solve this issue. This paper defines trusting the contents of a website using a new algorithm. To overcome trust issue, we have devised a way in which we can match and rank the web resources provided by a website using recommender systems and semantic web. At the end of this paper, we will see how much confident the algorithm introduced to the websites.

Keywords- Trusted websites, Content trusted scale, websites evaluation, websites ranking, recommender system matching and ranking, semantic web.

I. INTRODUCTION

The Internet consists of billions of websites, and web resources. Many of those webpages may have trusted contents while others can have untrusted content. As many of webpages do not provide quality content, from the user point of view, it is a complicated decision to trust whether the targeted webpage has a trusted content or no. This paper is exploring the possibilities of introducing a trusted content principle and the possibility of issuing a content trusted certificate in order to facilitate users to find quality and secure content. The concept works as follows: considering that W_1, W_2, ….. , W_n, are webpages, the objective of this paper is to find the relation between these webpages in the Trusted Content Scale (TCS). The relation is based on trusted contents.

We want to build the scale TCS which help us to find the TCS value for W_x compared with the TCS value of W_n. A variable the TCS value of Wx might be 0 to 10 based on its trusted content. If it is Zero (0) then its trust score is low or very bad and if the TCS value is Ten (10), so it is perfect. The TCS values for W_1, W_2, …. , W_n, can be evaluated using different evaluation methods which we are going to discuss in this paper.

In the current situation, we rely on web resources as it has an easy access to all kind of data. In addition, we do not ignore faulty information in such conditions so trust worthy content becomes our goal to be achieved.

The other problem for content security we face nowadays is social media. Social media plays an important role in our life as we are all connected to our loved ones and friends. At the same time, information shared on social media is sometime wrong, but people trust it without bothering to check whether that information is correct or not. Social media provides the freedom for everyone to post almost anything, so we cannot control illogical information shared by lot of people on daily basis. In order to overcome this issue we have to find a method to rank the trust level of the content even if it is provided by the most popular sites.

Some methods have been used earlier to overcome trust issues. The main and most important method used was ranking websites. The more the users use and browse the website the more trust level it gains. This method got lot of importance but it failed because of different reasons. For example, some websites are best resource to gather information, there are millions of topics discussed and posted on that website. There might be some pages with in those sites that share false information. We consider this as the main problem that we focus on in our research. We want to highlight the false information in every website. We will use our novel method, which works as follows.

The trust score is ranked from 0 to 10 where 10 stands for fully trusted site and zero stands for untrusted sites. After marking a site as trusted or

untrusted depending on its score we issue a certificate and use a logo to show it on the original website.
Next step is identifying who ranks the site and how. This is a very important step because it defines the score of the website. The algorithm reads the sites and only evaluators are allowed to provide rankings on basis of quality of contents the site provides. Designated experts are allowed to give scores to the site, so the common factor to evaluate the site is only trustworthiness. Based on this factor we are able to rank the content score from 0 to 10.
According to our technique, the main aspects used to judge the score is as follows:

Topic considered

Context and criticality of the need for information

Association by other trusted resources (i.e. citations)

Expertise of the user Perceived bias of source

Agreement with other resources

Precise and specific content

Likelihood of content being correct given what is known

Time of creation of the content

By applying the above criteria, we can achieve our objective to provide quality and trusted information to the users. As we checked that this method could be applied without spending lot of money, we consider that this method could represent a compulsory step for betterment of trusted resources on internet.

II. LITERATURE REVIEW

Not enough work has directly addressed the correctness of the contents of a website. However, Miguel et al. [1] presents a study on information security concerning mobile collaborative learning and how protection plays a major role in the design and execution processes of e-Learning in general and collaborative learning in particular. In order to advance collaborative learning security when adopting mobile learning models we have to examine new security approaches, which may be applied in this context. This has been done by giving an impression of safe learning management systems, examining which are the most significant factors to consider, and joining this method to the specific aspects of mobile collaborative learning. Learning is one of the applications that can be used by a large population of users and requires high degree of trust and correctness. It is also one of the applications widely used by mobile information systems users. In another research carried out by Jin [2], a new definition of mobile learning was illustrated and the features of mobile learning were introduced. This resulted in designing a framework for a specific mobile learning system. After that, nine function modules of the mobile learning system were mentioned. In addition to that, Yi [3] portrays the problem of mobile learning platforms as being heterogeneous and conducts a study on combining mobile learning with eLML. ELML, is an e-Lesson developed by University of Zurich that marks up language based on XML framework. It is autonomous on any special learning platform and it abides by SCORM standard. Besides, it is dominant, transplantable, and has the ability to join with other e-learning platforms. Yi basically introduced the method of designing and executing a mobile learning system based on eLML framework. It helps in finding a new solution to merge eLML mobile learning materials with current e-learning platforms available. Based on the introduction of mobile learning definition and characteristics, Li and Qiu [4] analyzed the characteristics of mobile learning theory and put forward a new mobile learning application model that combines the feature of learning targets with user demand. This research can provide new ideas for our content trusted scale to enhance the mobile learning theory. Jiugen et al. [4] discusses the background of the mobile learning, the theoretical foundation, and the practical applications. He discussed the possibilities of mobile learning in the future by putting into practice the following modes of learning: regularity of forgetting based refresher learning, Pleasant Education based Game Learning, discussion based cooperative learning and search engines based inquiry learning.
Within the context of social networking, Li et al. [5] addresses the question of how social networks like Facebook, twitter etc. affect real world just-in-time decisions and the importance of trusted information system. The question is critical due to the generality of mobile devices in our just-in-time decisions, the way we are connected to our social networks across time and space at various scales. The question is approached through true real world simulations where our mobile system has been deployed in a restaurant. The people's dish choices have been captured in a just-in-time manner. The collected data is used to show "how discrete choices are affected by the presence of physical and virtual social network information on our brains". Results show that virtual social network information has almost two orders of magnitude weaker influence than physical information on people's just-in-time choices. The modality in which our social information is presented affects people's decisions in the dimensions of time, price, and taste. Online Social Networks have become a basic medium for humans to manage their social life using mobile information systems. However, the structure of ego networks in such virtual environments has not been investigated yet. Arnaboldi et al. [6] contribute to fill this gap by analyzing a large data set of Facebook relationships. The data has been filtered to obtain the frequency of contact of the relationships, and the paper checked by

using different clustering techniques. According to Deng [7], the structural evolution of large online social networks has been considered. The results shown by theoretical analysis and simulation prove that, the evolution of the social networks consists with scale-free and small-world properties.. For some large networks from the Stanford repository and for a sample of social networks provided by Sistemi Territorial with high scarcity (80% and above) tests show that the algorithm consistently runs between one and two orders of magnitude faster than the current state of the art exact algorithms. In addition, the aim of the study conducted by Tobi et al. [8] was to investigate the use of online social networks and mobile information systems and its associations with university students' physiological, psychological, and social health. The study results found that the majority of the students participated in the Facebook (98%, n=417) and nearly half of the students spent two or three hours daily on the online social networking sites i.e. (49%, n=208). The students' participation for online social networks was significantly associated with the general health of students (r=0.26, p<; 0.05) and psychological health as (r=0.25, p<; 0.05). However, the findings show that online social networks did not affect the user's social health like satisfaction with social life, personal relationship, and social support.

III. METHODOLOGY

In this research, we propose an engine that can issue a certificate that is posted on a website. The certificate will have a number of a nine hexadecimal number. This number is posted with the logo of the issuer (the service provider, or the organization that is going to investigate the web-content and issue the certificate). This logo will show to the user the level of trust that can expected from the webpage. The level of trust could vary from 0 to 10. Where 0 is, the webpage was examined, but found un-trusted, and 10, is 100% trusted content.

This research was performed taking into consideration the following aspects:

a. How do I know if a webpage is a content-trusted webpage?
b. Can we build a platform for any organization to issue a content-trusted certificate to a webpage?
c. How can an entity (organization) investigate a webpage content? How the content is going to be examined and who is going to examine the content?
d. If the organization main webpage is trusted, does this imply that all the sub-WebPages are trusted?
e. Can someone use a fake certificate?
f. How to protect the logo?
g. How this mechanism may have an impact on the mobile information systems.

3.1 Sample collections:

To check the efficiency of our proposed system, we performed the following experiments:
We selected more than 400 WebPages randomly, as shown below:
100 Webpages for different websites W_1,W_2,W_99, belong to set #1 so Set # 1 contained {W_0, W_1, W_2, ... W_99 }
Similarly set # 2 contained {W_100, W_101, W_102, ... W_199 }
Where each webpage in set # 2 belongs to main page in set # 1, and each webpage in set # 1 has at least one webpage in set # 2.
Set # 3 contained {W_200, W_201, W_202,... W_299 }
And each element in set # 3 has to have an element in set # 2.
Consequently, it will have a main element in set # 1. This is happening in an arbitrary way. However, it is not necessary that each element in set # 2 has an element in set # 3. As a result, we may have a website with minimum number of WebPages = 2 and maximum number of WebPages = 7.
The WebPages were selected from 3 different languages
Set # 4 (languages) = {English, French, Arabic}
Set # 5 (Domain 1) = {mobile network, social network}
Set # 6 (Domain 2) = {learning}
Then we have asked the investigator to voluntarily evaluate the correctness of a webpage W_0, W_1, W_2, ... W_399 }
The investigators are in set # 7 {I0, I1, ... I49}.
Any investigator may rate 1 webpage or 10 WebPages, he or she may give a rate of 0, 1, 2, 3, 4, ….., 10 where 0= un-trusted and 10= very trusted content. His or her rate is subjective and based on his or her pinion.
In the experiment the webpage might be rated only once by the trusted investigators. Figure (1) shows the hierarchy architecture of the webpages in the experiment.

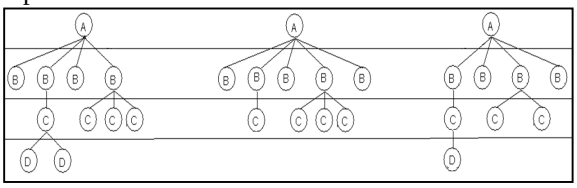


Figure (1): The hierarchy architecture of the webpages in the experiment.

3.2 Research design

Once the data collection is achieved, the webpage is sent to the appropriate investigator based on its language and domain. As shown in Figure (2), the investigator will rate each webpage. Then the cumulative rate is calculated based on equation (1).
Cr=(∑▒〖webpage rate〗)/(number of webpages) (1)
Cr=(∑_(i=1)^n▒〖wi rate〗)/(n+1) (2)
where Cr is the final rate given to the website.

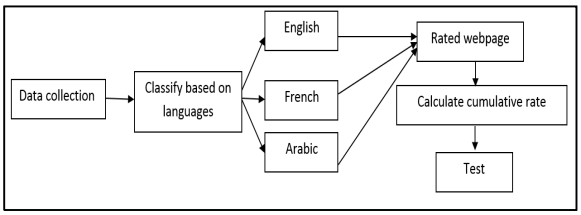


Figure (2): The ranking engine

The webpage W0 to W399 were proposed to the students in the testing phase. Then the students have to give their feedback on the rate of the WebPages.

3.3 Validity of the data

The students were informed to participate voluntarily, and they were cross-checked after they entered their rates. They were asked to rate the sites according to their knowledge and quality of content these web pages were providing. Then we showed them our ratings for those sites they rated and for most of them were an agreement between the two ratings. Consequently, our method is working totally fine as we are checking site content with the help of experts so they never make mistake to rank the sites. We can conclude that CTS system is totally working according to our proposed idea and people agree with our ranking system.

RESULTS

Figure (3) shows the result when we ask the users about their opinion in having the content trusted scale. 70.1% of the participants in the study said that it is very important to have a system to rate the webpage, so they can have some information about the trust level in a particular webpage. 25.4% suggested that it is very important for some websites, but not for all websites, whereas only 4.5% said it is not that important.

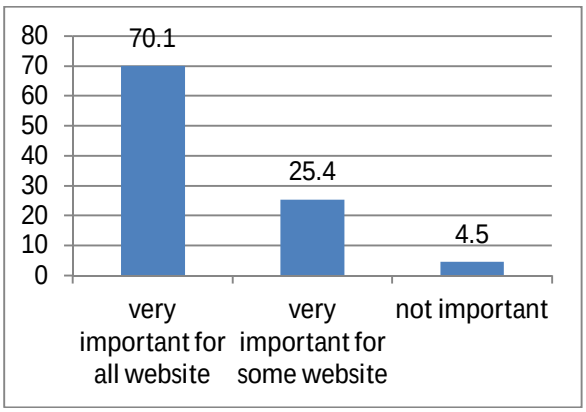


Figure (3): Opinion about the CTS

Then when we asked the users, do you feel confident if the CTS logo is exist? Their answers are plotted in Figure (4). As we can see in the Figure, 82.3% showed that they feel confident, 14.7% they feel little confident, and 3% only said they do not feel much confident.

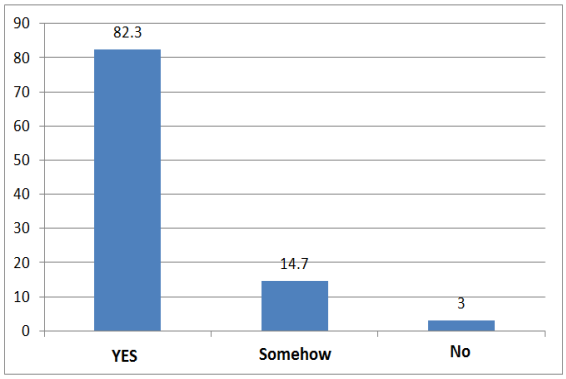


Figure (4): Confidence about the CTS logo

When we asked the users if the CTS logo encourage you to re-visit the website in the future. The answers are shown in Figure (5). 98.3% of the participants answered "YES" the logo encourage them to re-visit the website again, whereas 1.7% said "NO".

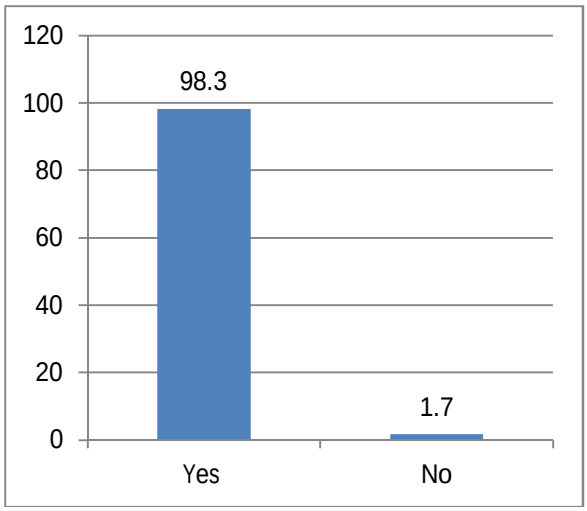


Figure (5): Re-visit in the future if the logo exist

Then we asked the users, if the rating is fair, “According to your background, do you agree with the rating given to the website you have visited? The answers were shown in Figure (6)

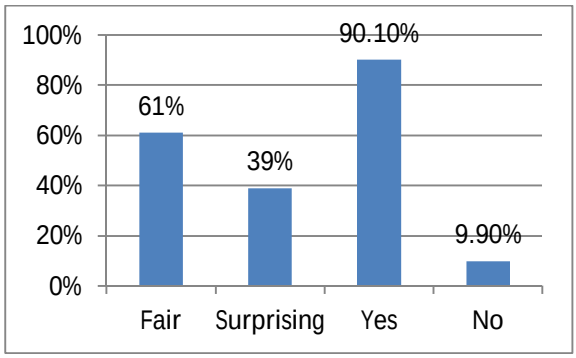


Figure (6): Do you think the rate is fair?

Figure (6) showed that61% of the participants agreed with the rating, and 39% where surprised with the rating. However, 90.1% said "YES" to the rating result, and 9.9% felt the rating were not as expected.

DISCUSSION

Figure (3) shows that when the users were asked whether they want CTS system on all the websites or not, more than 70% of the users said that they really want this system to be implemented on each and every site no matter whether it is popular or it is just a simple blog with very few visitors. 25% people said yes it should be implemented but only on limited number of sites as they want this system for only those sites which are not too much important. According to them low page rank sites need this system most. The other 4% of the users said that they do not need it. Consequently, from above results, we can conclude that most of people want this system to be introduced so they get better and trusted content.

In Figure (4), we can see whether the people feel comfortable or not when they see CTS logo on a site they visit. Almost 82% of the users reported that if they see a CTS logo on a site, it makes them feel comfortable because they think that, if the site is showing CTS logo, then it will be sharing trusted content only. The other 14% people said that they feel somewhat comfortable, but they cannot trust 100% because they think still wrong information can be shared by the site. The other 3% said they do not feel any comfortable. We can see from results that lot of people wants CTS system to be implemented, as they trust CTS checked sites. So CTS system will have a large impact on the users and most of them trust it.

Figure (5) shows the results of the question asked whether they would re-visit the site with CTS logo if they need. Almost 98% of the users said yes. This shows that people trust the CTS system. If a site owner wants to improve visibility to his website then he can do so by having a good rate on his site. In this way, lot of people will try to make their sites according to CTS requirements and they will remove wrong and false information published on their sites. This will increase competition between website owners and they will try their best to score 10 out of 10. Hence, our system will make the Internet a trusted content environment.

Figure (6) shows the results when people were asked whether they agree with the ratings given to a particular site. Almost 90% of the users said they agree with the ratings and other 9% said they do not agree. This result shows that CTS system works according to the requirements, as it was unbiased. The ratings provided to sites were correct depending upon their quality of content they provided. We can say that our system is satisfactory as people trust on our ratings system and they believe in it. Most of people were satisfied with ratings that we provided to those sites.

From the above results, we can conclude that CTS system will be the best evaluation system to eliminate wrong information shared by lot of websites. The next thing is the trust level we gained from people is satisfactory, as people liked those sites where our logo was shown by site owner and they trusted the content of site as it was marked as safe according to our system. The next important aspect that we noticed was that webmasters were keen to check their site's score, as most of people want good and quality information shared by their site. They were willing to change their sites if they had a low score. It is positive point, as it will create competition between websites and they will try their best to get good rankings and more visitors. Now, many websites are providing false information, but when our system will come in action, then there will be only trustable content on each site, so people will not need to evaluate whether they are reading correct or wrong content.

CONCLUSION

CTS can be explained as an evaluation tool and according to the context. In many cases, low quality sites provide information where most of people trust it and high quality sites provide information but many people do not agree with it. So content trust only depends upon the basis of information shared not based on site popularity.

Trust plays important role. No matter what information we get, we want it to be trustworthy. Many techniques were defined to obtain trust on Internet resources like ranking the websites on basis of page rank. This CTS will have a large impact on information systems as we saw in the simulation and in the survey. The information systems' users are interested in trusted contents.

Evaluators of the websites work hard to send websites to experts in every field to check the trust level of the site and the content on their own whether the content is up to standards or not. If the content is good then they will rank the content according to our scale. If content scores 100% results then it will be given trust score of 10 which will be maximum score similarly if a site scores 0 it means it is totally providing wrong information to its readers so we have to mark this site as un trusted site.

Once this step is achieved, our auto-engine will issue the certificates to all the sites according to their score. In this way, all sites which are providing wrong information will be eliminated. The system we are providing is a low cost and has high efficiency. Hence we can control the websites to share trust worthy content so everyone can use that information.

This paper present the first content trusted scale to trust a website based on a trusted evaluator, for a mobile information system innovation. The real problem arises when lot of people are posting some information according to their own understandings and many people start trusting that content as it was published on the Internet so it is perfectly right. We know that if the content is written by a person who belongs to a particular field with enough references the content, then we can trust it. So we created a

system called CTS where we will organize and manage partially the traffic of the Internet. Then we will rank all websites and their content according to our system. In this way we will classify the sites as trusted or untrusted depending on their score which they obtained according to our system.

In this way, we can eliminate lot of websites and there will be few websites on internet but all of them will provide real information. Then we will move towards social media sites to rank them according to our system as lot of people share and read from social media instead of websites.

Before our system content trust was measured basis of popularity [9].The more people visit a site, the more sites linking it, the better page rank that site will have and hence it was marked as trusted site. This is the point where old methods and techniques fail. We needed a system that will not only check the popularity of site but also the content it shares with its readers. In this way we can distinguish between trustworthy bloggers and others providing wrong information.

REFERENCES

[1] Miguel, J., Caballé, S., Prieto, J. (2013). Information Security in Support for Mobile Collaborative Learning. Third International Workshop on Adaptive Learning via Interactive, Collaborative and Emotional Approaches (ALICE 2013). Taichung, Taiwan. July 3-5, 2013. In proceedings of the Seventh International Conference on Complex, Intelligent, and Software Intensive Systems, pp. 379-385. IEEE Computer Society. ISBN: 978-0-7695-4992-7/13.

[2] Jin, Y.: 'Research of One Mobile Learning System', in Editor (Ed.) (Eds.): 'Book Research of One Mobile Learning System' (2009, edn.).

[3] Li, H. and Qiu, F., "Analyzing theory characteristic and studying application mode about mobile learning", Electrical and Control Engineering (ICECE), 2011 International Conference, pp. 6327-6330, 2011.

[4] Y, Jiugen, X, Ruonan and W, Jianmin, "Applying Research of Mobile Learning Mode in Teaching," Information Technology and Applications (IFITA), 2010 International Forum, Volume: 3, Page(s): 417-420, 2010

[5] Lippman, A., et al., "The Impacts of Just-in-Time Social Networks on People's Choices in the Real World", Privacy, security, risk and trust (Passat), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (socialcom), pp. 9-18, 2011.

[7] Deng, J., et al., "Study on evolution model and simulation based on social networks", Natural Computation (ICNC), 2012 Eighth International Conference, pp. 1238-1241, 2012.

[8] Tobi, S.N.M., et al., "The use of online social networking and quality of life", Technology, Informatics, Management, Engineering, and Environment (TIME-E), 2013 International Conference, pp. 131-135, 2013.

[9] Brin, S., and Page, L. (1998). "The anatomy of a large-scale hyper textual web search engine." In: Proceedings of the 7th International World Wide Web Conference

★★★