Identifying the Websites that Maintain Operational Standards through Obligation Links to Website-Standards Approval Body

Srihari Rao Nidamanuru[a], Rajesh Tiwari[b], Bhaskar Koriginja[c], Jincy Denny[d], J Ramesh Babu[e]

[a,c,d] Computer Science and Engineering, Bharat Institute of Engineering and Technology, Hyderabad, India
[b] Professor in Computer Science and Engineering CMR Engineering College Hyderabad, India
[e] Computer Science and Engineering BV Raju Institute of Technology Hyderabad, India
[a] n.sriharirao@biet.ac.in, [b] drrajeshtiwari20@gmail.com, [c] k.bhaskar@biet.ac.in, [d] jincy@biet.ac.in, [e] ramesh.jonnala@gmail.com

Abstract: Website is a very interesting phenomenon on the Internet. It attracts almost everyone as it can be viewed and used just by sitting in one place and exploring the many facets of information. That is the reason every institute whether for profit or non-profit establishes an online channel in the form of a website to manifest its significance both online and offline. Unfortunately, today the very purpose of the www on the Internet is abused by so many cybercriminals and their accomplices. By our many observations and website browsing and navigation experiences, we felt the need to assess any website and to approve/disapprove the special identification given to any website that meets certain operational standards. In this paper, we proposed a policy and a method to be adopted by Website-Standards Approval Body (WSAB) in order to assess the websites for approval/disapproval. This method is based on improving a parameter called Value-Identifier (VI) that is computed on the basis of certain important factors which are applied onto any website. We followed an empirical approach to assess the effectiveness of our method by following the operational standards to a set of 4 websites known to us. The larger and sufficient VI value is demonstrated by two of these operational-standard qualifying websites against the other two operational-standards non-qualifying websites.

1.   Introduction

Website is a very interesting phenomenon on the Internet. It grabs attention of almost every person as it can be an excellent online interface to the rest of the virtual world. This online facility can be availed just by anybody sitting in a place with a desktop/laptop computer plus an Internet connection and clicking/keying to view/explore many facets of information. That is the reason every institute whether for profit or not-for-profit establishes an online channel in the form of a website to manifest its significance and impact both online and offline. Woefully, today the very goal of the www on the Internet is abused by so many cybercriminals and their accomplices. Anybody with enough observation and website browsing/ navigation experiences, can feel the need to have an assurance of a standard body who will summarize and put before the website users about the safety and security features of a website. Standards are meant to achieve a level of quality in a process to produce a product or provide a service. In this paper, we focused on the Website-standards not directly on the W3C/Web standards. By Website-standards we mean, all the things that help the website to provide the quality information/service to its users in a convenient and time efficient manner.

2.   Motivation

The standard body should shoulder the responsibility to prevent cybercrimes that are generally associated with some of the websites. If the standard body assesses any website, it can approve/disapprove the special identification given to a website that meets/does not meet certain operational standards. In this paper, we proposed a policy and a method to be adopted by Website-Standards Approval Body (WSAB) in order to assess the websites for approval/disapproval. This method is based on improving a parameter called Value-Identifier (VI) that is computed on the basis of certain important factors which are applied onto any website. We followed an empirical approach to assess the effectiveness of our method based on the proposed operational standards on a set of 4 websites known to us. The differences in the VI values and the significance of the superior VI values can be demonstrated by applying our proposed policy and method on two of the selected quality websites against the other two selected non-quality websites.

A. How Website-Standards Approval Body (WSAB) is distinguished from W3C?

The World Wide Web Consortium (W3C) is an international community that develops open standards to ensure the long-term growth of the Web. W3C standards (Web Standards) define an Open Web Platform for application development that has the unprecedented potential to enable developers to build rich interactive experiences, powered by vast data stores that are available on any device. The Open Web Platform is the

collection of open (royalty-free) technologies which enables the Web. Using the Open Web Platform, everyone has the right to implement a software component of the Web without requiring any approvals or waiving license fees. On the other hand, WSAB is a large body consisting of website analysts who can accurately assess the quality of any website and has authority to approve/disapprove that website w.r.t its operational standards. The approval to a website by WSAB can impact largely on the amount of user/customer base to it.

B. Operational-Standards Identification

A website should not only be useful but also, it should be safe and secure to use. Currently, almost all the websites announce the safety and security features to its users but, there is no official body approving the same features for any website. That is, no website is being given any special identification for maintaining its standards and promising safety and security towards its users by any Standards Approval Body. Our proposal to give special identification to a website is based on the fact that any website should maintain the standards when it is operational for it to be considered by anyone (inside or outside its institution or organization or business) as useful, safe and secure to its users.

C. Obligation Link

Obligation is an act or course of action to which a person is morally or legally bound. Obligation also means something by which a person is bound or *obliged* to do certain things. Obligation Link is the requirement for a website-owner to undergo the evaluation process w.r.t his website by an organized body consisting of website analysts who can judge any website very well against a set of characteristics. All the approval aspirant website owners must approach the Website-Standards Approval Body (WSAB) and go through a scrutiny process with their website under consideration. If the WSAB stamps their approval for the website after adhering to the proposed policy and applying the method for operational-standard verification proposed in this paper, that website is considered to be Real Website (RW) otherwise it is Non-Real Website (NRW). Obligation Link is present in between the Website owner(s) and the WSAB and is for the above course of action through which the owner(s) of the website must undergo.

D. Quality Website (QW)

Any website under consideration can be treated as a Quality Website (QW) on the Internet if it satisfies most of the following properties:

1. It provides relevant, useful content and a good user experience.
2. If thought and time is invested into it to be useful to the users.
3. It clearly states its full postal address and phone number.
4. It is possible to speak to a real person associated with it.
5. It is transparent about its history and objectives.
6. It is professionally designed.
7. If there is consistency in style and presentation (fonts, colors, alignments, etc.).
8. Its text is clear and easy to read.
9. It is simple and logical to navigate.
10. It is free from broken links and 404 errors.
11. Its content is free from spelling, grammatical and factual errors.
12. It is free from manipulative SEO tricks.
13. The domain name is representative of the business's name.
14. It is free from viruses and malware.
15. It has a clear topical focus.
16. The content provides comprehensive coverage of its topic.
17. The content is written by a verifiable expert or enthusiast.
18. It is free from duplicate/overlapping/redundant content.
19. Its content is trustworthy/reliable/unbiased.
20. It is regularly updated.

21. Its content is unique/valuable/engaging.

22. Its content is of the quality that you might expect to see in a printed magazine/journal/book.

23. It provides substantial value when compared to competitor websites.

24. It is cited on, and linked to from, other high quality websites.

25. It references and cites other high quality websites.

26. There is a fair balance between providing value and marketing.

27. There is a fair balance between content and ads.

28. Ads are clearly distinguishable from content.

29. It avoids unnecessary registration and data collection.

30. There is a privacy policy when the registration is required or data is collected.

31. It loads properly across all web browsers and screen resolutions.

32. Pages load in less than 3 seconds.

33. The average visitor visits more than 1 page.

34. The average visit duration is more than 1 minute.

3.   An Instance wherein the Website Users can be Misled

Organic search, aka natural search, refers to unpaid search results. In contrast to paid search results as in pay-per-click advertising, which are populated via an auction system, organic search results are based on relevance to the user's search query, links and domain authority and other organic ranking factors. SEO stands for Search Engine Optimization, which is the practice of increasing the quantity and quality of traffic to your website through organic search engine results. Generally, a website owner opts to increase the organic search results for his website enormously to have dominant ranking position. In this process, it is likely that he can opt for SEO tricks to achieve higher organic traffic and increase in sales through internet marketing. It is possible for any website user to be misled when the website does not contain valuable things but, merely applies SEO tricks to attract and trick the website users.

4.   Proposed Policy and Mechanism

If a website besides satisfying the characteristics of a Quality Website (QW), gets approval and a standard logo from the Website-Standards Approval Body after an extensive and thorough scrutiny by it, then that website can be considered to be a Real Website (RW). Otherwise, that website is a Non-Real Website (NRW). We propose a mechanism and a policy for a website to be worthy for consideration as a RW in this paper. The proposed policy and the mechanism together can act as guidelines for the WSAB in order to judge any website as either RW or NRW and therefore approve or disapprove that website. An ill-designed website (potential NRW) affects its users by violating many of the above characteristics needed for a RW and therefore does not deserve approval by the WSAB. WSAB analyst assesses a website w.r.t all the above mentioned 34 characteristics as in column 2. If a particular characteristic is satisfied, the analyst will assign integer value '1' for that website in the column corresponding to it, otherwise he will assign integer value '0' to it. After that, VI is calculated by dividing the total count with 1's in that column by the number of all non-NA values. Then based on the value for VI, a website may be categorized under RW or NRW. If the value of VI ≥ 0.75, then that website may be classified as RW, otherwise it is classified as NRW.

A RW gives the users the following advantages:

1. Ensures the safety of money and time

2. Provides accurate and desirable information

3. Ensures Transparency

4. Ensures Accountability

5. Increases Respect to the User

6. Enables the Defusion of Fake Websites

7. Binds Culprit Website Owners to Law

8. Increases the Credibility of Online Transactions over

a Website

9. Provides Value for Customer's Time Investment

5.   Results from Empirical Study

We have collected the assessment of four different websites made by two analysts (Analyst1 and Analyst2) to yield four sets of averaged VI values. We mark the website under RW, if the yielded VI value is more than or equal to 0.75, otherwise, we mark it under NRW. Following this policy, we assigned RW for the first two websites and NRW for the other two websites as in table 1. The calculated VI values by each analyst for a specific website are depicted on the histogram separately as shown in figure 1. Set IDs are the identifiers for a website which was analyzed by two website analysts w.r.t the 34 parameters given in table 1. For example, Set ID value 1 indicates that 1 is the identifier for the Website1.

6.   Conclusions

With the help of day to day observations and website browsing and navigation experiences, we recognized the need to assess any website and to approve/disapprove the special identification given to any website that meets/does not meet certain operational standards. In this work, we proposed a policy and a method to be adopted by Website-Standards Approval Body (WSAB) in order to assess the websites for approval/disapproval. This method is based on a parameter called Value-Identifier (VI) that is computed on the basis of certain important factors which can be applied onto any website. For doing this, we conducted an empirical study to assess the effectiveness of our proposed method by verifying operational standards to a set of 4 websites known to us. The higher VI value is demonstrated by two of these operational-standards qualifying websites against the other two operational-standards non-qualifying websites. Therefore, we recommend our method to the potential WSAB as a remedy and recovery approach against the losses incurred by customers who have been deceived by the websites that exploit various tricks to cheat the website users/customers.
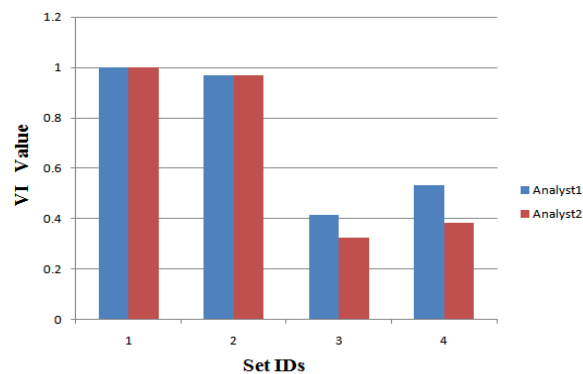

Figure 1. Values of VI for each Website by each Analyst are Shown in the Histogram Separately

7.   Acknowledgment

References

1. https://www.seoexpert.co.uk/high-quality-websites/#:~:text=In%20 short%2C%20a%20high%20quality,as%20a%20high%20quality%20one.
2. https://www.wordstream.com/organic-search
3. Shanshan Qi, Crystal Ip, Rosanna Leung, Rob Law, "A New Framework on Website Evaluation", International Conference on E-Business and EGovernment, 2010,pg.78-81.
4. Dan Zhao, Chunhui Tan, Yutao Zhang, "Evaluating the Enterprise Website Credibility from the Aspect of Online Consumers", International Conference on Management of e-Commerce and e-Government, 2009.
5. WensongHu, MinZhu, Van Zhang, Yingpei Wu, "Studies on Websites Reliability", International Conference on Optics, Photonics and Energy Engineering, 2010.

6. Jason Koepke, Siddharth Kaza, Ahmed Abbasi, "Exploratory experiments to identify fake websites by using features from the network stack", IEEE International Conference on Intelligence and Security Informatics,11-14 June 2012, pg.126-128.

7. Abbasi. A and Chen. H, "A Comparison of Tools for Detecting Fake Websites", IEEE Computer (42:10), 2009, pp. 78-86.

8. Ntoulas, "Detecting Spam Web Pages through Content Analysis", Proc. 15th Int'l Conf. World Wide Web (WWW 06), pp. 83-92, 2006.

9. Maclnnes, D. Musgrave and J. Laska, "Electronic Commerce Fraud: Towards an Understanding of the Phenomenon", Proc. Hawaii Int'l Conf. Systems Sciences (HICSS 05), pp. 181.1-181.11, 2005.

Table 1. Properties for Verification Vs. Values (1/0/NA) for each Property in Column2

S.No	Properties for Verification	A1's V* for WS1	A2's V* for WS1	A1's V* for WS2	A2's V* for WS2	A1's V* for WS3	A2's V* for WS3	A1's V* for WS4	A2's V* for WS4
1	It provides relevant, useful content and a good user experience	1	1	1	1	1	1	1	1
2	If thought and time is invested into it to be useful to the users	1	1	1	1	1	1	1	1
3	It clearly states its full postal address and phone number	1	1	0	0	0	0	1	1
4	It is possible to speak to a real person associated with it	1	1	NA	NA	0	0	0	0
5	It is transparent about its history and objectives	1	1	1	1	0	0	0	0
6	It is professionally designed	1	1	1	1	0	0	1	0
7	If there is consistency in style and presentation (fonts, colors, alignments, etc.)	1	1	1	1	0	0	1	0
8	Its text is clear and easy to read	1	1	1	1	0	0	1	1
9	It is simple and logical to navigate	1	1	1	1	1	1	0	0
10	It is free from broken links and 404 errors	1	1	1	1	1	1	1	1
11	Its content is free from spelling, grammatical and factual errors	1	1	1	1	0	0	1	0
12	It is free from manipulative SEO tricks (hidden text, cloaking, doorway pages, etc.)	1	1	1	1	1	1	1	1
13	The domain name is representative of the business's name	1	1	1	1	1	1	1	1
14	It is free from viruses and malware	1	1	1	1	1	1	1	1
15	It has a clear topical focus	1	1	1	1	0	0	0	0
16	The content provides comprehensive coverage of its topic	1	1	1	1	1	0	0	0
17	The content is written by a verifiable expert or enthusiast	1	1	1	1	0	0	0	0
18	It is free from duplicate/overlapping/redundant content	1	1	1	1	0	0	0	0
19	Its content is trustworthy/reliable/unbiased	1	1	1	1	0	0	0	0
20	It is regularly updated	1	1	1	1	0	0	1	1
21	Its content is unique/valuable/engaging	1	1	1	1	0	0	1	1
22	Its content is of the quality that you might expect to see in a printed magazine/journal/book	1	1	1	1	0	0	1	1
23	It provides substantial value when compared to competitor websites	1	1	1	1	0	0	0	0
24	It is cited on, and linked to from, other high quality websites	1	1	1	1	0	0	0	0
25	It references and cites other high quality websites	1	1	1	1	1	1	1	1
26	There is a fair balance between providing value and marketing	1	1	1	1	0	0	0	0
27	There is a fair balance between content and ads	1	1	1	1	0	0	0	0

28	Ads are clearly distinguishable from content	1	1	1	1	0	0	0	0
29	It avoids unnecessary registration and data collection	1	1	1	1	1	1	1	0
30	There is a privacy policy when the registration is required or data is collected	1	1	1	1	1	1	0	0
31	It loads properly across all web browsers and screen resolutions	1	1	1	1	1	1	0	0
32	Pages load in less than 3 seconds	1	1	1	1	0	0	0	0
33	The average visitor visits more than 1 page	1	1	1	1	1	0	1	1
34	The average visit duration is more than 1 minute	1	1	1	1	1	0	1	0
	Total V* values	34	34	32	32	14	11	18	13
	Value Identifier(VI)	34/34	34/34	32/33	32/33	14/34	11/34	18/34	13/34
	Average VI	1		0.969		0.367		0.456	
	RW/NRW?	RW		RW		NRW		NRW	

where V* = Value for a Property in Column2, *NA stands for Not Applicable

Value Identifier(VI) = (Total V* values)/(34 - Number of NAs)and

Average VI= (A1's VI for Wn+ A2's VI for Wn)/2