# FORM INPUT VALIDATION
## *An Empirical Study on Irish Corporate Websites*

Mary Levis, Markus Helfert
*Dublin City University, Dublin 9, Ireland*

Malcolm Brady
*Dublin City University, Dublin 9, Ireland*

Keywords: Information System, Website Design Quality, Form Input Validation.

Abstract: The information maintained about products, services and customers is a most valuable organisational asset. Therefore, it is important for successful electronic business to have high quality websites. A website must however, do more than just look attractive it must be usable and present useful, usable information. Usability essentially means that the website is intuitive and allows visitors to find what they are looking for quickly and without effort. This means careful consideration of the structure of information and navigational design. According to the Open Web Applications Security Project, invalidated input is one of the top ten critical web-application security vulnerabilities. We empirically tested Twenty one Irish Corporate Website. The findings suggested that one of the biggest problems is that many failed to use mechanisms to validate even the basic user data input at the source of collection to validate user input in order to ensure reliability and therefore potentially resulted in a database full of useless information.

## 1 INTRODUCTION

The World Wide Web (WWW) is the largest available distributed dynamic repository of information, and has undergone massive and rapid growth since its inception. There are over 2,060,000 users in Ireland alone. Over the last seven years (2000 - 2007), Internet usage in Ireland has grown by 162.8%; in United Kingdom by 144.2% in Europe by 221.5% and Worldwide by 244.7% (www.internetworldstats.com).

Based on these facts, the Internet creates a greater need for businesses to design better websites in order to stay competitive and increase revenue. The Website's homepage should be a marketing tool designed as a 'billboard' for the organisation. The design is critical in capturing the viewer's attention and interest (Mandel, 2002) and should represent the company in a meaningful and positive light. Most often in the drive to make the website look appealing from a visual perspective other factors are often ignored, such as validation and security, which leads to poor user experience and data quality problems.

Data is deemed of high quality if it '*correctly represents the real-world construct to which it refers so that products or decisions can be made'* (Pike, Barnes**,** 1996**)**. One can probably find as many definitions for quality on the web as there are papers on quality. There are however, a number of theoretical frameworks for understanding data quality. Redman and Orr have presented cybernetic models of information quality. The cybernetic view considers organizations as made up of closely interacting feedback systems that link quality of information to how it is used, in a feedback cycle where the actions of each system is continuously modified by the actions, changes and outputs of the others (Beckford, 2005; Orr, 1998; Redmond, 1995). Wang and Strong proposed a data quality framework that includes the categories of intrinsic data quality, accessibility data quality, contextual data quality and representational data quality outlined in table 1.

Levis M., Helfert M. and Brady M. (2008).
FORM INPUT VALIDATION - An Empirical Study on Irish Corporate Websites.
In *Proceedings of the Tenth International Conference on Enterprise Information Systems - HCI*, pages 148-153
DOI: 10.5220/0001681601480153

Table 1: IQ Dimensions (Source: Wang, 1998).

| DQ Category | DQ Dimensions |
|---|---|
| Intrinsic DQ | Accuracy, Objectivity, Believability, Reputation |
| Accessibility DQ | Accessibility, Access Security |
| Contextual DQ | Relevancy, Value Added, Timeliness, Completeness, Amount of Data |
| Representational DQ | Interpretability, Ease of understanding, Concise Representation, Consistent Representation |

Quality of web sites may be linked to such criteria as timeliness, ease of navigation, ease of access and presentation of information. From the customer's perspective usability is the most important quality of a Web application (Fraternali, 1999). The root cause that leads to web application problems is the poor approach to web design. Several techniques exist to evaluate the quality of websites for example link checkers, accessibility checkers and code validation, to name a few. In practical terms this required validating a site against a series of checkpoints that included: checking that legal and regulatory guidelines were adhered to pages conformed to Web-Accessibility standard; missing page titles; browser compatibility; user feedback mechanisms; applications were functioning correctly (e.g. online forms are validated for input etc.); clear ordering of information; broken links; page download speeds and ease of navigation.

Online interactivity is a valuable way of improving the quality of business web sites and web designers should be aware of how design affects the quality of the web site and the image of the organisation. One important factor for a web site being successful is speed. If the web site is unresponsive, with long response times after clicking on links, the visitors will not come again.

For the purpose of this study we conducted an empirical study of twenty one finalists in a recent website quality technology award that included (3) Charity/Not for Profit organisations; (7) Large Quoted Companies, (2) Small Quoted Companies and (9) Statutory and Unquoted Companies. The aim of this study was to examine these websites for *Technical* quality issues specifically form input validation. The rest of the paper is organized as follows: Section 2 outlines what makes a quality website, Section 3 shows our methodology, and Section 4 gives a brief summary and conclusion.

# 2 A QUALITY WEB SITE

A good website must include safeguards against failure and provide simple, user friendly data entry and validation processes. Information is regarded as an important factor impacting organizations. From the literature reviewed a universal definition of information quality is difficult to achieve (Orr, 1998; Stylianou, 2000; Strong, Lee and Wang, 1997; Wang 1998; Bugajski, Grossman and Tang 2005; Kumar, Ballou and Ballou 1998; Olson, 2003). According to Mandel (2002), *'Technically, information that meets all the requirements is quality information'.*

## 2.1 Links

Good websites have a rich and intuitive link structure. A link going to the Customer Service should be named 'Customer Service' and the surfer looking for Customer Service information will know this link goes to the page they want. Therefore, 'click here' should never be used as a link. A good web designer will think clearly about how each piece of data links up with the rest of the content on the website and will organise the links accordingly**.**

## 2.2 Navigation

Without a clear navigation system, viewers can become disoriented. Therefore it is good practice to have a simple, consistent navigation structure throughout the site. Users use browsers to navigate through WebPages and hyperlinks are distinguished from normal text within a page by its colour. When the page pointed to by a hyperlink has been 'visited' browsers will inform the users by changing the link's colour (Tauscher and Greenberg, 1997).

## 2.3 Input Validation

According to the Open Web Applications Security Project (OWASP, 2007) invalidated input is in the top ten critical web application security vulnerabilities. For this reason, input validation is an important part of creating a robust technological system and securing web applications. One solution is to provide input data validation at the data collection point before the form is submitted. Incorrect data validation can lead to data corruption. Input validation should be performed on all incoming data ensuring the information system stores clean, correct and useful data. Trusting users to enter the correct data is never a safe assumption.

Examples of invalid data are: text entered into a numeric field, numeric data entered into a text field, or a percentage entered into a currency field. Table 2 provides an example set of checks that could be performed to ensure the incoming data is valid before data is processed or used.

Table 2: Example Validation Checks.

| Validation checks | Description |
|---|---|
| Character Set | Ensure data only contain characters you expect |
| Data Format | Ensure structure of data is consistent with what is expected |
| Range Check | Data lies within specific range of values |
| Presence Check | No missing / empty fields |
| Consistency Check | If title is 'Mr' then Gender is 'Male' |

## 3 RESEARCH APPROACH

We conducted an empirical study on a recent website technology award using the full data set of twenty one finalists that included (3) Charity/Not for Profit organisations; (7) Large Quoted Companies, (2) Small Quoted Companies and (9) Statutory and Unquoted Companies. The aim of this study was to evaluate sites against a series of checkpoints. The principle used was based on the same criteria used to evaluate the participants in the 2006 award. However, for the purpose of this study we used a partial set of the criteria as outlined in table 3.

Table 3: Partial set of criteria.

| Validation Criteria |
|---|
| Contrast colours support readability & understanding |
| Professional appearance having no horizontal scroll bars |
| No links labelled 'click here' |
| What is clicked on is also title of page jumped to |
| Links to other pages & back to home page are functional and relevant |
| Help features are available and easy to access |
| Visited links change colour |
| Site map provided |
| Interactive form validated for input |
| Mailto parameters are set correctly |
| Web address is simply a case of adding .com .org or .ie to company name |
| Useful search engine is provided |
| Site search is provided |
| Frequently Asked Questions page provided |
| Data Protection and Privacy |
| Users can see Feedback other users have provided |
| Data protection act, privacy & business ethics |

### 3.1 Findings and Analysis

Table 4 shows the number of companies who defaulted and the number who adhered to the selected criteria.

Table 4: Criteria selected for Website evaluation.

| Validation Criteria | Defaulted | Adhered |
|---|---|---|
| Contrast colours support readability & understanding | 8 | 13 |
| Professional appearance- No horizontal scroll bars | 2 | 19 |
| Avoided using 'click here' | 12 | 9 |
| What clicked on was also the title of page jumped to | 1 | 20 |
| Links to other pages & to home page functional & relevant | 6 | 15 |
| Help features available & easy to access | 11 | 10 |
| Visited links changed colour | 16 | 5 |
| Site map provided | 6 | 15 |
| Interactive forms validated for input | 17 | 4 |
| Mailto parameters set correctly | 6 | 15 |
| Web address was simply a case of adding .com .org or .i.e. to company name | 4 | 17 |
| Useful search engine provided | 19 | 2 |
| Site search provided | 7 | 14 |
| Frequently Asked Question Page provides | 10 | 11 |
| Data Protection & Privacy | 12 | 9 |
| Feedback forms | 21 | 0 |

Twelve websites did not include a link to their data protection and privacy policy. Ten companies did not have a FAQ link and 11 did not have help features available and easy to access. Seven sites did not have the mailto parameters set correctly. Fifteen sites had fully functional and relevant links to other pages and back to the homepage. Thirteen sites promoted contrast colours supporting readability and understanding and 19 had a professional feel and appearance. Twenty sites adhered to the criteria of having the title of page jumped also as the label of the link connecting to it.

Figure 1 depicts the results from the Friendly URLs (Uniform Resource Locator). Seventeen (81%) had good structured semantic URLs, made up of the actual name of the specific company where we could guess the URL by simply adding .com, .org or .ie to the company name. For example a company named '*Jitnu*' had a URL http://www.jitnu.ie or http://www.jitnu.com as the web addresses which convey meaning and structure. Only 19% defaulted on these criteria having a URL such as http://www.jitnu.ie/?id=478 instead of http://www.jitnu.ie/services

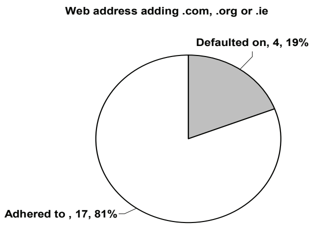

Figure 1: Results of the Friendly URL's criteria.

Figure 2 shows that 16 (76%) used the same link colour for visited and unvisited pages and did not support a convention that users expect. Good practice is to let viewers see their navigation path history (i.e. pages they have already visited) by displaying links to '*visited pages'* in a different colour.

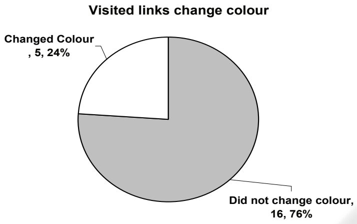

Figure 2: Visited Pages changed link colour.

Sitemaps are particularly beneficial when users cannot access all areas of a website through the browsing interface. From analysis of our findings in figure 3 we show that six websites (29%) did not provide a site map.

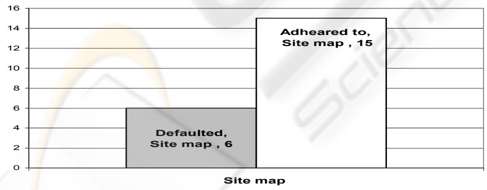

Figure 3: Site Map.

In figure 4 we show that although adding a search function on a web site helps visitors to quickly find information they need, seven (33%) failed to provide a comprehensive site search or search interface. While creating a good navigation system will be sufficient help for many people, it won't meet the needs of everyone.

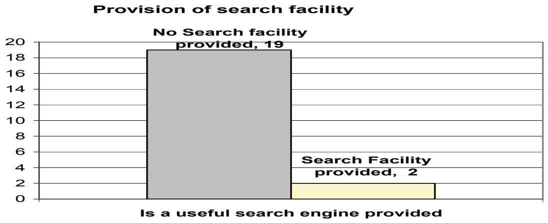

Figure 4: Search Facility.

Figure 5 shows the results of checking user-entered email addresses for valid input. An email address should contain one and only one (@) and contain at least one (.). There should be no spaces or extra (@). There must be at least one (.) after the (@) for an email address to be valid. Many sites implemented some form of email address validation but did so incorrectly. For example they correctly rejected jitnu.eircon.net and jitnu@eircom@net as invalid email addresses; however, they incorrectly accepted 'jitnu.eircom@net', as a valid email address. While they correctly checked for the presence of the (@) and the (.), they did not however check the order in which the (@) and the (.) appeared in the email address. We found that 17 (81%) had no validation process on email addresses while only 4, (19%) shown in figure 5 had complete validation.

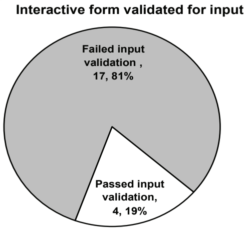

Figure 5: Email Validation.

Figure 6 shows that 12 (57%) were careless about their link text quality by using *'Click Here'* which does not give any indication of the content on the linked page, while nine (43%) used meaningful link text which identified the target of the links.

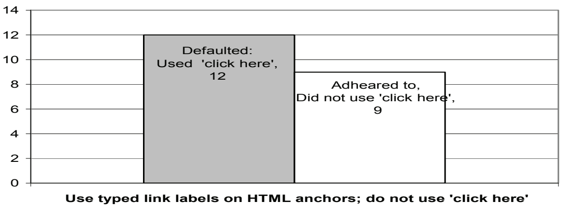

Figure 6: Click Here Anchors.

Figure 7 shows the number of companies that adhered to the quality criteria set out for this review and figure 8, the number that defaulted in the above criteria. It can be seen that 19 sites had a professional appearance and 20 sites used the page title of the page linked to as an anchor. However only 2 sites provided a site search option and 4 sites had complete validation on email addresses.

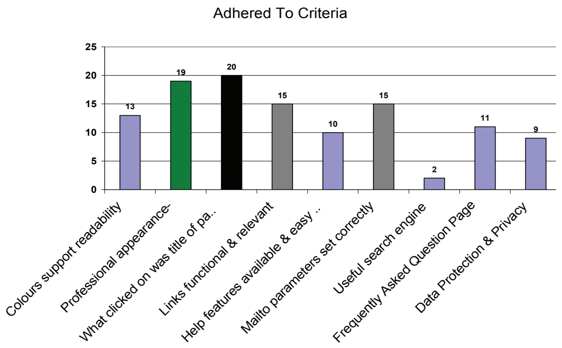

Figure 7: Adhered to the selected criteria.

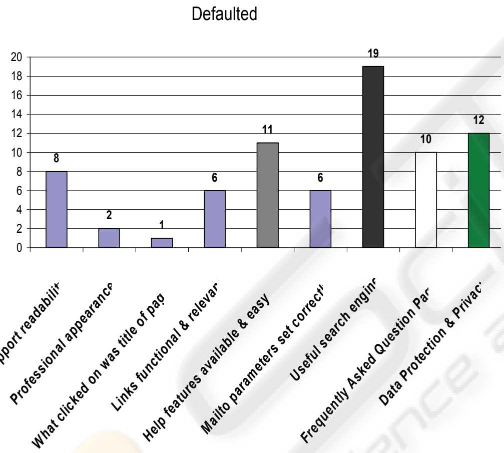

Figure 8: Defaulted on the selected criteria.

Figure 9 shows that six sites under review did not have their ‘mailto’ parameter set correctly to facilitate the user with easy feedback option and none of the 21 sites provided an option for the users to view the feedback provided by other users.

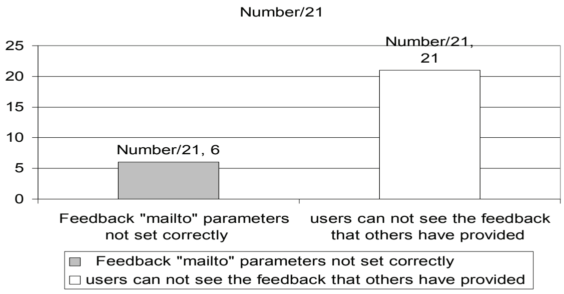

Figure 9: Feedback option and accessibility.

From our analysis we found that all sites had good page load speed between a minimum 0.5 seconds, a maximum of 1.5 seconds and an average load speed of .89 seconds.

## 4 SUMMARY / CONCLUSIONS

Today's Internet user expects to experience personalized interaction with websites. If the company fails to deliver they run the risk of loosing a potential customer forever. An important aspect of creating interactive Web forms to collect information from users is to be able to check that the information entered is valid therefore; information submitted through these forms should be extensively validated.

Not validating input is one of the biggest mistakes that Web-application developers make. This can lead to database corruption. Our results show that one of the biggest problems, with online forms on the web is that many corporate companies failed to validate user input to ensure reliability and potentially result in a database full of useless information. Validation could be performed using client script where errors are detected when the form is submitted to the server and if errors are found the submission of the form to the server is cancelled and all errors displayed to the user. This allows the user to correct their input before re submitting the form. Our study highlights that to date many web applications have not used mechanisms to validate even the very basic data input at the source of collection.

Although the majority of web vulnerabilities are easy to avoid many web developers are unfortunately not very security-aware. A company database needs to be of reliable quality in order to be usable. A simple check whether a web site conforms to the very basic standards could have been done using the W3C HTML validation service, which is free to use. Web developers need to become aware and trained in Information Quality Management principles, and especially in the information quality dimensions as outlined in table 1.

Given that the sites under review were among large quoted companies, small quoted companies, charities and not for profit, statutory and unquoted organisations, some of which had been recognised for excellence in financial reporting, it was surprising to find that 81% of the sites under examination failed on the basic input validation. One Hundred percent of large and small quoted companies failed in their email input validation while 67% of charities/not for profit organisations and statutory and unquoted organisations failed to validate emails. No less than 90% of all organisations under review failed to provide a useful search engine and only 71 % provided a site map. However, 67% provided a site search facility and 81% had friendly URL's and most sites had good design layout that was consistent throughout making it easier for the user to navigate.

Many problems could be eliminated by checking for letters (alphabet entries only); numbers ( numeric entries only); a valid range of values; a valid date input; and valid email addresses. Keeping in mind that a user could enter a valid e-mail address that does not actually exist it is imperative that some sort of activation process needs to be done in order to confirm a valid and correct email address.

## REFERENCES

Beckford John, 2nd edition, Quality, Rutledge Taylor and Frances Group, London and New York (2005)

Bugajski Joseph, Grossman Robert L., Tang Zhao, An event based framework for improving information quality that integrates baseline models, casual models and formal models, IQIS 2005 ACM 1-59593-160-0/5/06. (2005)

Fraternali, P., Tools and Approaches for Developing Data-Intensive Web Applications: A Survey, ACM Computing Surveys, vol.31, No.3, (1999)

Internet_World_Statistics http://www.internetworldstats.com

Kumar Giri, Ballou Tayi, Ballou Donald, P., Guest editors, Examining data Quality, Communications of the ACM, vol. 41, No 2, pp 54-57. (1998)

Mandel Theo, Quality Technical Information: Paving the Way for UsableW3C Web Content Accessibility Guidelines 1.0, \\http://www.w3.org/tr/wai-webcontent/

Olson Jack E Data Quality: The Accuracy Dimension, Morgan Kaufmann, ISBN 1558608915. (2003)

Open Web Application Security Project, http://umn.dl.sourceforge.net/sourceforge/owasp/OWASPTopTen2004.pdf

Orr Ken, Data Quality and Systems, Communications of the ACM, vol. 41, No 2, pp 66-71, (1998)

Pike R.J., Barnes R *TQM in Action: a practical approach to continuous performance improvement,* 1996, Springer, ISBN 0412715309

Print and Web Interface Design, ACM Journal of Computer Documentation, vol. 26, No. 3. (2002)

Redmond, Thomas C, Improve Data Quality for Competitive Advantage, Sloan Management Review, vol 36, no 2, pp. 99-107 (1995)

Strong, Dianne M., Lee Yang W., Wang Richard Y., Data Quality in Context Communications of the ACM, vol. 40, No 5, pp 103-109. (1997)

Stylianou Antonis C., Kumar Ram L, An integrative framework for IS Quality management, Communications of the ACM, vol. 43, No 9, pp 99-104. (2000)

Tauscher, L., Greenberg, S., How people revisit web pages: Empirical findings and implication for the design of history systems, International Journal of Human-Computer Studies, 47, 97-137 (1997)

Wang Richard Y., and Strong, D.M. *Beyond accuracy: what data quality means to data consumers*, Journal of Management Information Systems 12, (4), pp 5–34. (1996)

Wang Richard Y., A product perspective on Total Data Quality Management, Communications of the ACM, vol.41, No.2, pp58-65. (1998)