EVALUATION OF THE USABILITY OF WEB-BASED APPLICATIONS

Nebojša D. Đorđević
Serbian Armed Forces, Army HQ, Niš, Republic of Serbia,
e-mail: djnebojsa@mts.rs,
ORCID iD: http://orcid.org/0000-0003-3730-0085

http://dx.doi.org/10.5937/vojtehg65-11319

FIELD: Computer Sciences, IT
ARTICLE TYPE: Professional Paper
ARTICLE LANGUAGE: English

Summary:
The paper emphasizes the importance of the usability of Web-based applications as an essential condition for attracting and retaining customers. At the beginning of the paper, a general classification of methods for usability evaluation is given in order to show different views of researchers on usability. In order to ensure a Web application lifetime, it is necessary to measure and evaluate many features that affect software usability. The paper gives a brief overview of the most commonly used methods for evaluating the usability of Web-based applications published in the last decade, chosen by the author. Since decision making is not an easy process, the conclusion gives recommendations to be specially considered when selecting a method.

Key words: methods, usability, Web-based applications.

Introduction

The emergence of the Internet contributes to the speedy development and massive use of Web-based applications. The specific properties of the Internet as a basic working and development environment for Web applications indicate that Web applications represent a rather specific software product. When the Web is concerned, however, the need for a positive user experience in interacting with the application is further emphasized, ie. satisfaction and comfort level of users in achieving the objectives of a Web site, as an essential condition for the retention of the user.

Given that Web applications have been developed considerably shorter than classic information systems, the evaluation of usability is often
skipped, and the reason for this lies in the fact that the application of certain methods takes time, expensive and sophisticated equipment and the participation of experts as evaluators. However, designers of Web applications are aware that the evaluation of usability can significantly affect the reduction of the development of Web applications if the usability problems are identified in the early stages of the life cycle. Therefore, the basic question of practitioners is to find the most efficient way of integrating Web usability evaluation into daily work.

Today, there are various methods for assessing usability. As a result, there is a question of choice of the most appropriate method for assessing the usability of a particular software product. The choice of an adequate method can significantly improve the efficiency of the evaluation process and usability of the software product. However, the choice of the appropriate method to be applied in the process of evaluating usability is not simple, since it depends not only on the software product type, but also on the development of the objectives of the project and the context of use. In fact, the choice of a method depends on various criteria, some of which are among the most important resources required to perform the method (time, money, the number of evaluators and their expertise, the number of users for testing, place and test equipment), the required level of objectivity and the possibility of applying development of Web applications in various stages. Using a combined approach can reduce the disadvantages of different usability methods and find a good compromise between the needs to implement high quality evaluation of Web usability, and the time and cost of execution.

Specifics of Web applications

Between Web applications and traditional software systems, there are some differences arising from the specific environment in which they are developed, maintained and used. In "cyberspace", the Internet and the Web remove restrictions of real distances in space allowing instant access to information regardless of how far users and servers might be away from each other. This quality of the Web provides numerous advantages of Web applications over traditional, desktop applications, and they include:

- **Global approach.** Web applications are published centralized in one place and the whole world can see them. Any user who has access to the Internet can access Web applications from a home computer.

- **Simultaneous work of a large number of users.** In general, traditional desktop applications are used by one user at a certain time,
while Web applications can be used by tens and hundreds of users simultaneously. Web applications are usually intended for large, diverse, remote user groups, which have many varied requirements and expectations in terms of national, religious and cultural feelings and standards, different levels of knowledge and a variety of platforms for the application use. This requires a greater need for security and privacy, and higher standards and performances of Web applications compared to desktop applications.

– **Ability to work on multiple platforms.** Most clients of Web applications are Web readers who play the role of a universal interface between the user and the system for displaying data of any format and can be run on any computer with the Internet access. Web applications use publicly accessible and free Web browsers and do not depend on the user’s software platform. Due to the fact that there are different Web browsers typical for different operating systems (Internet Explorer, Mozilla Firefox, Apple Safari, Google Chrome, Opera, etc.), and that all these readers largely consist of HTML and JavaScript standards, Web applications relying on HTML clients typically support different operating systems.

– **Low cost compared to the average number of users.** Most Internet components are free for end users, which also applies to Web applications. Organizations that have a need for a Web application can reduce the cost of its purchase and maintenance because employees can set up and use the Web application at home, at work or under field conditions.

– **Ease of use by end users.** Web applications are designed for a broader audience so they are simple to use, similarly to regular Web sites. The ease of use of Web applications encourages public participation, but obliges Web developers to customize this application to users who have no previous knowledge.

– **Centralized upgrade.** The process of upgrading Web applications is faster and easier, because the changes implemented are centralized in one place, so changing a program code on the Web server becomes immediately visible to all users.

– **Different purpose of use.** Unlike desktop applications whose use is limited to a certain number of users, Web applications can be used by a wide number of users, for a variety of purposes of business and personal nature.

– **Web applications can fail for many different reasons.** The timescales for the development of Web applications are significantly shorter and this influences the choice of methods and techniques for their
development. Non-linear navigation and unpredictable user behavior and the environment in which Web apps work (limited bandwidth or availability of Web servers) may affect the user experience (Murgesan, 2008, pp.7-32). The consequences of failure and dissatisfaction of users of Web applications can be much more serious (more expensive) than in the case of traditional desktop applications.

The above features indicate that Web applications are fairly specific software products. For these reasons, many researchers in the area of quality Web applications in their works (Bublione et al, 2002) (Becker, Olsina, 2010), (Olsina, Molina, 2008, pp.385-420), (Lew, Olsina, 2011, pp.214-229) indicate that the existing quality models listed in the relevant ISO/IEC standards are not suitable to describe the quality of Web applications.

In the era of hyper-production of complex and sophisticated Web software, usability is crucial for the acceptance of Web applications and is the key quality factor that determines their success or further destiny.

Along with the growth in popularity of Web applications, the attention paid to the evaluation of their use in all phases of the life cycle has also increased.

Methods for evaluating the usability of Web applications

Just as there are a lot of different approaches and definitions of quality, there are numerous methods for its evaluation. The methods can be qualitative or quantitative, automatic, semi-automatic and manual, ranging from easy to difficult to use, etc. Most of the available methods originated from Human-Computer Interaction (HCI) and are primarily intended to evaluate the quality of traditional software products. Although they are used to identify usability problems of traditional graphic user interfaces, today they can be equally successfully applied to a variety of Web applications.

There are plenty of general quality models tailored specifically for Web applications, but the efforts of researchers over the past decade have given a number of models of software quality Web applications, oriented to a specific domain (Đorđević, 2017, pp.513-529). However, this chapter will show the most commonly used methods from the last decade, developed to evaluate the quality of Web applications.
WAMMI

WAMMI (Website Analysis and Measurement Inventory) is a Web analysis service that measures and analyzes the experience of real Web site users to help them achieve the digital goal (Muylle et al, 2004, pp.543-560).
WAMMI is a measuring tool which:
- measures the Web site user experience based on the reactions of visitors.
- compares the site in relation to other Web sites in the international standardized database.
- generates objective data for management and a convenient digital report easy to read.
- analyzes qualitative comments of users and their reactions to the site.
- interprets quantitative and qualitative data to determine what to improve and how much to invest.
WAMMI is a research Web site and an analytical service developed using psychometric techniques with a data reliability rating between 0.90 and 0.93.
It is based on international software standards and expertise obtained from the assessment of software usability. It is used in the public sector (e-government) and business sectors such as banking, finance, travel, telecom and IT, and web sites for electronic commerce (e-commerce). WAMMI is often used for international studies and is available in most European languages.
Statistical methods are applied to select 20 statements that summarize the essence of the experience of site visitors from a large number of questions. Each question is a vital aspect of the user experience and they are all required to cover the entire spectrum of customer experience. All questions cover specific topics, such as attractiveness, control, efficiency, helpfulness and ease of learning. Visitors fill out questionnaires and give their answers. A digital report is generated at the end of the probationary period. Visitor experiences are measured through questions to compare their expectations with what they found on the site. A few additional questions help to analyze detailed information on the type of visitors to the site, the reasons why they visited it and how they think it can be improved.
WAMMI has a unique approach because it compares the satisfaction of site visitors who estimate the values of the reference database, which contains data from more than 320 selected analyses. This allows to
compare the site being estimated with the other one. Other questionnaires can only give the number of visitors who assessed the site.

The whole process takes a few minutes only. When enough users answered (between 40 and 200 users), a digital report is received within two working days and the whole assessment process usually takes no longer than three weeks.

The most important element of the report is the profile of the site, which contains five sub-scales (Figure 1): Attractiveness, Controllability, Efficiency, Helpfulness and Learnability) and there is a general assessment of Global Usability.

![Figure 1 – Graph of the results about the website profile](image)

If the Web site at any scale is estimated above average (50) according to the database, it is given as a green bar and extends above the 50 line. But if the site achieved a score below the average on the scale, this is indicated by a red bar extending down from the 50 line. The average score is 50; below 30 or above 70 means the site is remarkable on this scale, while a perfect score is 100.

The standard deviation expresses a degree of variation in the data. For this type of data, a reasonable value of the standard deviation is 20:00. The greater the number of respondents agreeing in their assessments of the Web Site, the smaller the standard deviation, and vice versa - if many respondents have different opinions, the standard deviation will be much higher. The standard deviation over 30 indicates that there are two or more
groups of subjects with very different opinions about the site usability. It is not uncommon that standard deviations vary in scales. This indicates that there are differences in the degree of agreement of respondents about these scales.

Other elements of the WAMMI report are:
- detailed analysis for each statement and setting priorities for the site aspects which need improvement.
- analysis of additional questions with fixed response categories.
- answers to the free text questions where visitors talk about things that are not specifically required by the WAMMI questions.
- profiles of individual visitors and a numerical summary of the WAMMI results.

**UWIS**

HCI (human-computer interaction) professionals generally discover perceptual and motor difficulties through problems based on skills and problems of rule-based consistency, while the true intentions of end users are identified by the questions based on knowledge (mental models) (Abdinnour-Helm et al, 2005, pp.341-364).

This shows that there is a need for a comprehensive methodology for measuring the usability of Web-based information systems, which will integrate measures of quality and usability. UWIS (Usability of Web-based Information Systems) is a methodology for assessing the usability and design of Web-based information systems that combines the size and quality of Web services and the usability of information systems (Oztekin et al, 2009, pp.2038-2050).

To assess the usability and quality of Web-based information systems, UWIS uses appropriate methods. This methodology applies the structured equation model SEM (Structural Equation Modeling) to establish a quantitative model for evaluating usability. UWIS integrates the established dimensions for measuring the quality of Web services with the appropriate lists of formulated questions, which is a modification of the ServQual model, expanded with the dimension of usability. To create a list of questions, UWIS uses ServQual and WebQual approaches to measure the quality, the principles of dialogue for the design of the user interface according to ISO 9241-10 (ISO, 1996) and Nielsen usability heuristics (Nielsen, 1994).

The UWIS methodology defines a quantitative model for measuring the dimensions of usability and introduces two latent variables called the usability index (Figure 2). In accordance with the definition of usability in
ISO 9241-11 (ISO, 1998), effectiveness, efficiency and satisfaction are high-level parameters that are grouped and aggregated in the index of usability 1 (UI₁). These dimensions are objective measures of usability and cannot be changed directly and consciously by the user interface designer.

The dimensions of the low level of usability are the following measures: reliability, integration of communication, navigation, controllability, assurance, responsiveness, and quality of information. They are collected through the UWIS methodology for forming usability index 2 (UI₂). Low level dimensions can be changed directly and usability can be improved by analysts and designers using the user interface.
To measure the connections and relationships between the indexes of usability UI1 UI2, classical statistical methods of multiple regression are used in a combination with the factor analysis.

The UWIS methodology may give a list of the most critical dimensions. Once they are repaired, the performance of the usability of Web-based information systems is expected to improve significantly (through efficiency, effectiveness and satisfaction) because there is a strong link between the rate of low and high level of usability. The correlation analysis was used to determine the numerical indicators of the strength and direction of the relationship between variables.

The main limitation of the UWIS methodology is that it does not provide a solution to measure the usability of Web-based information systems if the lists of dimensions are not linearly associated with the usability index. This flaw stems from the basic principles of the SEM quantitative method applied by UWIS. In such cases, it is necessary to use sophisticated analytical techniques such as genetic algorithms, neural networks and vector regression to explain the relationship between the non-linear lists of the dimensions and the index of usability.

**WebQual**

Similarly to the previous one, this method is based solely on the view of the end user who is considered to be the ultimate judge of quality. This qualitative method is classified as a test method, because it uses the mechanisms of the questionnaire which uses a set of 36 statements to assess 12 factors of quality of Web applications, classified in four categories of the highest level: usefulness, usability, fun and building relationships (Figure 3). It is mainly designed to assess whether the user will visit the site again. The TAM (Technology Acceptance Model) is used as a theoretical basis for defining the criteria based on which the user will decide to do so.
While filling out the questionnaire, the site user expresses his/her agreement and disagreement with a statement on the seven-point Likert scale ranging from "strongly disagree" to "agree completely".

Using competent assessors of the quality of Web applications, WebQual provides a quite reliable method of assessment. Selected quality factors provide good opportunities for the establishment of an area that is "the most problematic", so that its improvement is a priority. When it comes to the evaluation of the quality of the site, this method has the best price/quality ratio, simply because users fill out a questionnaire for free, and the information site owners receive is extremely valuable and relevant.

A disadvantage of this methodology is that (Loiacono et al, 2002) the method of analyzing the data obtained from the questionnaire is not clearly
defined. The Likert scale is a scale of ordinal values, where the responses are classified by ranks, but the intervals between them cannot be considered equal. This means that the method of calculation of the mean (and standard deviation) cannot be used for the analysis of ordinal variables. Appropriate techniques of descriptive and deductive statistics differ for ordinal (i.e. qualitative) and lapse (or quantitative) variables, and if WebQual users use wrong statistical techniques, it could easily happen that they draw wrong conclusions from the collected data thus "fixing" something that is not necessary to repair and neglecting actual shortcomings of the Web site.

**WEF**

WEF (Website Evaluation Framework) (Zhou, 2009) is a quantitative methodology that supports the thesis that customer satisfaction is more important than anything else, which means that it neglects other important user roles (e.g. development and maintenance).

The main objective of this methodology is the evaluation of any Web site, regardless of the domain, type or programming/script languages. The advantages of this methodology are its universality and simplicity. It allows each owner or administrator of the Web site to automatically and easily check whether the site is in accordance with the rules of good practice, without the need to have technical and/or domain knowledge.

Although this concept is a great idea in the field of software quality assurance, the relevance and practical usability of an evaluation template like this is questionable.

The WEF Quality Model consists of five quality characteristics (Zhou, 2009): aesthetics, ease of use, multimedia, rich content, and reputation. Only two of them, aesthetics and ease of use, are divided into subcharacteristics (Figure 3). Other measures are direct indicators of quality. The importance of individual quality factors is determined by the assigned numbers that represent fixed weighting factors.

The evaluation approach of this methodology is from the bottom to the top, which means that the values of the most basic factors of quality indicators are measured first and then summed up by an aggregation formula into high order quality factors (subcharacteristics and characteristics).
Aesthetic Evaluation

1. Images
   1.1.1 Definite Image Size
   1.1.2 One Larger Image in One Page
   1.1.3 Image ALT
   1.1.4 Image Link

1. Page’s Resolution & Resizable Table
   1.2.1 Sizable Table
   1.2.2 Optimize the Page Resolution

1.3 Color
   1.3.1 Using Multiple Colors
   1.3.2 Using Sage Color
   1.3.3 Limitations of Colors for Color Blindness People

1.4 Emphasis
   1.4.1 Underline of Text

Ease of Use

2.1 Consistency
   2.1.1 CSS Attributes
   2.2 Navigation
      2.2.1 Frames Validity

Multimedia

3.1 Plug-in Support
3.2 Attributes of Multimedia Components
3.3 One Media in One Page
3.4 Using Thumbnails

Rich Content

4.1 Bulletin Boards
4.2 Information Guide
4.3 Search Engine
4.4 Avoiding Auto-refresh

Reputation

5.1 Customer Feedback
5.2 Web Traffic
5.3 Domain Name
5.4 Information Publicity

After completing all evaluation steps, the site is ranked into one out of five categories, in accordance with the key shown in Figure 4 (Zhou, 2009).

<table>
<thead>
<tr>
<th>category</th>
<th>range</th>
</tr>
</thead>
<tbody>
<tr>
<td>★★★★★</td>
<td>0.9 — 1</td>
</tr>
<tr>
<td>★★★★</td>
<td>0.7 — 0.89</td>
</tr>
<tr>
<td>★★★</td>
<td>0.5 — 0.69</td>
</tr>
<tr>
<td>★★</td>
<td>0.3 — 0.49</td>
</tr>
<tr>
<td>★</td>
<td>0.1 — 0.29</td>
</tr>
</tbody>
</table>

In the end, it seems that it is too superficial to be used for a serious and comprehensive analysis and perhaps it could be used to evaluate the
quality of simple websites. It can be said that the simplicity of this methodology is its greatest strength but also its greatest disadvantage.

*WebQEM*

To provide methods and techniques, Olsina and Rossi (Olsina, Rossi, 2002, pp.20-29) presented the WebQEM (*Web Quality Evaluation Method*) method based on the C-INCAMI framework for quality measurement and evaluation.

The C-INCAMI methodology (Becker, Olsina, 2010) (*from Contextual Information-Need Concept model, Attribute, and Metric Indicator*) is a comprehensive and well-developed framework for the implementation of projects of measuring and evaluating quality; it prescribes a set of activities, their inputs and outputs, roles, interdependences, etc. which ensures the consistency and reproducibility of the measurement and evaluation process and its results.

The C-INCAMI framework consists of six basic activities:
1. Definition of non-functional requirements;
2. Planning of measurements;
3. Execution of measurements;
4. Planning of evaluation;
5. Execution of evaluation;
6. Analysis of the results and making recommendations.

Using WebQEM for the evaluation of Web sites and applications supports the efforts being made to meet the demands for quality in new Web development projects, as well as in those already operating. It also helps identify missing properties or poorly implemented requirements, such as interface design, or problems with navigation, accessibility, search systems, content, reliability and performances (Olsina, Rossi, 2002, pp.20-29).

The steps of the WebQEM process are grouped into four main technical phases:
1. Definition and specification of requirements for quality;
2. Elementary measurement and evaluation (planning and realization);
3. Global evaluation (planning and evaluation);

During the phase of defining and specifying quality requirements, the goals of the evaluation and the user point of view (the role) are specified. Then, a quality model is selected and it may be defined in the appropriate ISO standard, with the addition of the attributes specific for a particular domain. The relative importance of these components for selected users is then identified as well as the required level of coverage.
User roles can be classified into three abstract categories: visitor, member of the development team and manager. These categories can be broken down into sub-categories. For example, the visitor category can be divided into the sub-categories of conventional and advanced visitors.

When the domains and product descriptions are defined, the objectives agreed upon and the user role (i.e. explicit and implicit customer needs) selected, the next step is to specify the necessary characteristics, sub-characteristics and attributes in the form of a tree of requirements. The result of this phase is the specification of quality requirements.

The phase of the elementary measurement and evaluation defines two main activities: design of elementary evaluation and implementation of elementary evaluation. In the design phase, all information about the selected metrics and indicators is recorded, in line with the conceptual scheme of the Metrics and Elementary Indicator.

The phase of global evaluation has two main stages: design and implementation of partial and global evaluation. In the design phase, the aggregation criterion is selected as well as the scoring model. These two parameters are intended to make the evaluation well structured, accurate and understandable. There are at least two types of models: those based on linear additive scoring models and those based on nonlinear multi-criteria scoring models. Both types use the weighting factor as a way of determining the relative importance of indicators.

Even if we ignore the rest of the C-INCAMI framework (whose integral part WebQEM is), and observe this method separately, we immediately see its good sides. Concise, yet flexible, the proposed model of quality, a well-defined process and a scoring preference method based on a mathematical model of weight exponents make it one of the best resources for quantitative expert evaluation of the quality of Web applications that professional and academic communities currently have to offer. In addition, WebQEM can be used in the early stages of the development of Web applications as efficiently as the operational Web application. This is a possibility that the above described two methodologies do not have.

WebQEM has its drawbacks, though, the biggest of which being a necessity of expert evaluators who possess the knowledge necessary for defining the requirements tree (Zhou, 2009) and a good knowledge of the domain in which the Web app works. Therefore, this method carries the risk that, during the evaluation of global quality, subjectivity cannot be completely avoided (Olsina, Rossi, 2002, pp.20-29). Manual and thorough evaluation would require a huge effort and a lot of time, which may pose a
potential problem. For this reason, Olsina et al created a tool called "C-INCAMI Tool" in order to facilitate the evaluation process and save time.

Conclusion

An important component of software is its evaluation. To ensure the required quality, it is necessary to measure many characteristics that allow the determination of software quality, where software quality metrics play a significant role. However, first it is necessary to define a model with a set of quality characteristics of software quality to be assessed. Of course, it is not possible to measure all the characteristics of quality in all possible cases.

However, separate measurements are not suitable for evaluating the overall usability because each metric is measured on its own scale and the results are difficult to compare. The interpretation of usability across multiple metrics becomes clumsy, heavy and unconvincing for decision-making, which represents a drawback of such an approach.

A difficult task for professionals, business managers and potential customers is to determine which product is superior regarding its usability when considering several attribute measures on different scales. One usability metric provides better assessment of usability and easier comparison of products than when individual metric components are considered. However, the existence and use of these methods indicate a need to present the complex usability structure in a form that can be manipulated. Therefore, it is expected that the combination of a number of known evaluation methods provides an easily applicable procedure of the comprehensive and objective evaluation of the usability of Web applications, thus enabling easy identification of problems in the design of interfaces as well as an efficient comparison of competing products or the same product in different stages of its life cycle.

Performing a summary usability metrics reduces the complexity of identifying the difference in the usability of competing products and facilitates decision-making. In this way it provides clear, understandable and unambiguous interpretation of the results and readily compares the results of the usability of competing products or one particular product after changes.
References


ОЦЕНКА УДОБСТВА ИСПОЛЬЗОВАНИЯ ВЕБ-ПРИЛОЖЕНИЙ

Небойша Д. Джорджеевич
Вооруженные Силы Республики Сербия, Управление сухопутных войск, г. Ниш, Республика Сербия

ОБЛАСТЬ: вычислительная техника и информатика,
исполнительные технологии
ВИД СТАТЬИ: профессиональная статья
ЯЗЫК СТАТЬИ: английский

Резюме:
В данной статье подчеркивается роль удобства использования Веб-приложений, так как оно привлекает и удерживает пользователей на сайте. В вводной части статьи приведена общая классификация оценочной методологии по удобству использования, с целью обзорного представления о различных точках зрения отдельных исследователей в вопросах по удобству использования. Для продления жизненного цикла Веб-приложений необходимо проводить оценку и анализ многих факторов, влияющих на удобность использования программного обеспечения. В данной статье приведен краткий обзор наиболее распространенных оценочных методов в области удобства использования Веб-приложений, представленных в ряде публикаций, посвященных данному вопросу особое внимание на протяжении последнего десятилетия. В заключении статьи, автор, учитывая тот факт, что у многих пользователей вопрос о выборе методологии вызывает затруднения, приводит обоснованные рекомендации по преимуществу отдельных методов.

Ключевые слова: методы, удобство использования, Веб-приложения.

ЕВАЛУАЦИЈА УПОТРЕБЉИВОСТИ ВЕБ АПЛИКАЦИЈА

Небојша Д. Ђорђевић
Војска Србије, Команда Копнене војске, Ниш, Република Србија

ОБЛАСТИ: рачунарство и информатика, информационе технологии
ВРСТА ЧЛАНКА: стручни чланак
ЈЕЗИК ЧЛАНКА: енглески
Сазетак:
У раду је наглашен значај употребљивости веб заснованих апликација као битан услов за привлачење и задржавање корисника. На почетку рада представљена је уопштена подела метода за евалуацију употребљивости, са намером да покаже различит поглед истраживача на употребљивост. Да би се обезбедио животни век веб апликације потребно је мерити и вредновати многе карактеристике које утичу на употребљивост софтвера. У раду су укратко приказане најчешће коришћене методе за евалуацију употребљивости веб заснованих апликација публикованих у последњој деценији. Имајући у виду да није лако одабрати праву методу, у закључку су наведене препоруке које треба посебно разматрати приликом њеног избора.
Кључне речи: методе, употребљивост, веб засноване апликације.