DOI: 10.15849/IJASCA.230720.09

Int. J. Advance Soft Compu. Appl, Vol. 15, No. 2, July 2023
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Usability Evaluation of Information Retrieval web-based systems using User Testing and SUS methods

Majed A. Alshamari

Department of Information Systems, College of Computer Sciences and Information Technology; King Faisal University, KSA
*Correspondence: smajed@kfu.edu.sa

Abstract

Unsurprisingly, the web has become a standard resource to access information. Universities and organizations have significantly switched their attention towards web-based information systems. It is challenging to design an Information Retrieval (IR) system which can search relevant information and ensure both high usability and user satisfaction. High usability results in reduced usage time, eliminated frustrations and increased satisfaction. Because the IR websites differ in technical characteristics and scope, it is also challenging to design and conduct the usability tests of these systems. Previous studies examining the usability of IR websites have researched a limited number of IR websites and/or based their analysis on a review of the search criteria and search results, and thes studies have addressed quality and accuracy of the search results. This paper provides a much-needed overview of popular academic IR websites from a usability perspective. We compared different design styles and layouts of popular IR websites. By employing usability testing and SUS techniques, the perceived difficulty has been analyzed, while users perform the given tasks. Our test results assert that no matter how accurate an IR system is, users might find it difficult or impossible to utilize if it is not usable.

Keywords**: *HCI; Usability Testing and Evaluation; Information Retrieval, SUS technique.***

1 Introduction

Mobile and smart gadgets (such as phones, tablets, and smart TVs) or specialized devices with internet connectivity have made it feasible for individuals of all ages to access digital information today. Thousands of new applications, covering a broad variety of categories and uses, are uploaded to the Apple App Store and the Google Play Store every day [1]. A recent analysis by Grand View Research, Inc. [2] predicts that the mobile health (mHealth) app market size would achive US $149 billion by 2028, growing at a CAGR of about 17% between 2018 and 2028. The mHealth app category is a sizable one in the app stores, and it has been gaining popularity over the last couple decades. Up to 34% of mobile phone users, some studies report, have one health app as a minimum installed on their mobile devices [3]. Smartphones and tablets have also accelerated the use of artificial intelligence (AI) in mobile applications that span healthcare systems, finance, and entertainment [4]. In this era of exponential technological

Received 3 March 2023; Accepted 6 May 2023

advancement, AI-powered mobile applications (apps) are used by users from all walks of life. Evidently, AI is progressively becoming a major part of people's everyday lives [5], [6].

Using AI-based healthcare apps is important to patients; it is thus important that they are beneficial and not harmful. AI systems should thus provide patient satisfaction across multiple healthcare environments and be effective and efficient [7]. Hence, by examining the usability of mHealth apps, we can uncover their issues, help redesign systems, spend less time and money, and improve user acceptance [8]. When defining the usability of a system, effectiveness, efficiency, learnability, ease of use, and user satisfaction are considered some of the most common attributes. An often-used analytical method for usability evaluation is cognitive walkthrough (CW) [9].

Lewis and Wharton developed CW to evaluate the usability of interfaces using theory-based evaluation [10]. It is employed to identify problems and generate proposals about their causes. Learning through CWs is aimed to simplify learning, especially through exploratory learning. In medical equipment evaluation, CW is used to evaluate depression screening models [11], nurse information systems [12], diabetes management systems [13], and other healthcare systems. It is advantageous to use CW in healthcare, as it can be used to identify important usability problems quite easily, quickly, and cheaply when real usability testing is not feasible [14].

The rest of the paper is organised as follows: Section 2 discusses previous studies that evaluated the AI-based system. Section 3 describes the detailed methodology of our evaluation process. Section 4 presents the qualitative and quantitative findings of our evaluation. Finally, Section 5 details our conclusions.

2 Related Work

Usability and aesthetics are crucial factors to consider while designing a website. Many studies cited by Palmer [3] demonstrate the correlation between aspects including download speed, page layout, information architecture, content kind and volume, personalization, responsiveness, and the performance of web-based portals in terms of usability. Sonderegger and Sauer [10] used a computer simulation of a mobile phone to demonstrate how the visual look of the phone might alter perception of its usage, performance, and beauty. Users of the more aesthetically pleasing gadget rated it higher in use and attractiveness, even though the two phones were otherwise equal. Moreover, the time spent doing activities was computed and said to be less than the spent time in the case of the appealing ones, underscoring the influence that an appealing design may have on a user experience as a usability component.

Even if they aren't directly connected to the process choice, structural aspects such as typeface and typography [11], color [12-14], images [11,15], or information and data placement on a given page [16,17] may contribute to satisfying usability. Providing a satisfying and memorable encounter for the user is one way to do this [11,18]. User perception by extension, desire to buy or showed a choice may be influenced by color [19]. Bonnardel et al. [19] shown that users' reactions to a system's color scheme can be predicted with high accuracy. Both color and images have the potential to grab the attention of web users [11,15], but improper management of then may render them useless. Laeng et al. [20] and Rik and Michel [21] found that visuals, namely images, are the most popular parts of a website. When it comes to the design of the site itself, the writers of [22] analyzed the impact of several factors on the usability of websites. According to the research conducted by Tichindelean et al. [11], there are some important structures and designs that might increase the usability of website landing pages. According to their findings, readers are more likely to focus on the content located at the page's top, middle, and right. Participants' focus is mostly on the text, as noted by Matzen et al. [23].

The usability of a website depends on a number of factors, including the names of each part, the placement of the pieces, and the size of the elements. Zlokazova et al. [24] conducted research in a similar field, looking at how the layout of an airline's website might affect user experience. Despite making a number of mistakes, the research found that the two-page format helped individuals to complete their tasks more quickly. Users' increased mental effort may explain this discovery.

Local and specific organization website layouts have been researched and analyzed in the near past as well. Gossen and Nurnberger [25] have conducted a survey of young users' behavior about IR websites. In their findings they listed the shortcomings of the website designs and suggested improvements. In another literature review, Zahabi et al. [26] surveyed medical data interface usability issues. They evaluated the safety techniques and proposed design improvement guidelines. Anupriya and Kaur [27] compared and contrasted eight different life insurance websites (LIC, 92 SBI Life, HDFC Life Insurance, and Max Life Insurance) based on trust, aesthetic appeal, innovation, information fit-to-task, personalized information, response speed, intuitive operations, and relative benefit. Jiuqiang aimed to measure the usability of the four software stores by testing effectiveness, efficiency, and satisfaction [28]. he concluded new design goals and design principles from his software usability test which spanned over ten participants. Similarly, Bataineh and others in [29] explored the usability of three Dubai e-government units' websites utilizing eye tracking technique. They focused on understanding the participants' navigation behaviors on the web page based on pre-arranged tasks in addition to factors that affect their information searching behavior. Kaysi and Topaloglu [30] compared the usability of an existing student information system with a new system developed after the usability improvements. They found that the second system's design outperformed the first in terms of usability metrics the including total task time, number of fixations, and the average length of each fixation. Similarly, in [31], Pawel and others published the results of a study that aimed to evaluate the usability of selected web services using eye-tracking and concurrent thinking aloud (CTA) techniques. Duygu and Kursat [32] examined usability of five Turkish e-government portals with nine participants using a combined method consisting of quantitative and qualitative approaches. Their study identified the usability problems encountered while using the government services, and then suggested improvements. In another recent study, Sauer et al. [33] rehearsed the impacts of operational conditions on usability testing. They did a direct comparison between the remote and traditional field testing and testing in a lab. Their research showed that there were important distinctions between field and lab-based usability assessment in the face of challenging operational situations (e.g., dual task demands, poor product usability).

As far as the research on IR websites is important, it has also attracted the attention of the researchers. Diane and Kelly [34] outlined the primary factors to consider when assessing an IR system's interactivity and proposed many experimental layouts and sampling techniques. They also discussed central instruments, as well as the validity and reliability concerns with regards to the measurements and the applied techniques. Savolainen and Kari [35] investigated the relevance criteria that the users establish for themselves when selecting hyperlinks in online searches. Talking aloud data from 9 self-directed Google searches was analyzed quantitatively and qualitatively. The findings revealed that although many different criteria were used, certain criteria, like specificity, tended to exert dominance. Kengeri et al. [36] studied the effects of popular digital library attributes on search and retrieval attempts. For the development of future digital libraries, they developed a taxonomy of features and shared user suggestions. Methods for evaluating the usability of digital libraries were created by Judy and Jeng [37]. Usability's many facets were covered, and the article established the close connection between usability's three primary factors: efficiency, effectiveness, and user happiness. Brenda et al. [38] reviewed major usability principles and explored the applications of formal usability testing to an existing library website. Kelly L and Diane H [39] investigated the use of criteria to assess relevant, partially relevant, and not-relevant documents. The importance of context and relevance for the usability testing of IR websites has been studied by many researchers [34,35,39–41]. Carol and Barry [40], for instance, explored a relevant assessment by allowing participants to provide their own search statements to an IR system. However, recent and comprehensive studies are limited in this area.

In addition, Khabsa and Giles [42] recently used Google Scholar and Microsoft Academic Search to study the coverage of IR websites. Thorpe and Lukes [43] examined 430 public library websites in Indiana. Thorpe and Lukes studied a total of 129 characteristics on each homepage, including those related to navigation, search, content, and Web 2.0. The results provided a profile of public library homepage design while revealing widespread patterns in that area. Anderson [44] evaluated various usability testing approaches in a user-centered context and relevance, observing searches and relevant evaluations in a formal usability laboratory. In addition to their prior expertise and knowledge, Kammerer and Gerjets [45] analyzed the effect of CTA instructions. The findings revealed that, in comparison to neutral instructions, participants' verbal comments on information quality and on structural characteristics of Web sites were much higher when prompted CTA instructions were used.

2.1 Research Gap and Goals

There are previous studies which have examined the usability of IR websites. However, they have either emphasized a limited number of IR websites [38,43,46] or have focused their research on the searching capabilities of IR websites for systematic review [7,47] and have addressed the quality and accuracy of the search results. Without a doubt, the capability of IR systems to retrieve results in an effective and efficient manner determines its suitability in systematic searches. We, however, argue that users might find these search results difficult to utilize if it is not usable.

On the other hand, as far as the research on usability testing has been increased, it has either focused on only one specific personality trait like extraversion [48], on one type of usability metrics like survey data [49], or on a specific IR website [43]. In the ever-changing domain of IR there are newer usability elements and layouts being employed. Various usability studies, such as Bucher and Schumacher [50], showed that the layout and design have a high influence on visual attention to information and memorization.

Motivated by these findings that high usability ensures high user satisfaction, the present research seeks to compare and analyze different popular IR websites and identify the design styles, layout and formats which increase the usability of an IR website. Specifically, this research provides a much-needed overview of popular academic IR websites from a usability perspective. Our research examines different IR websites and analyzes their unique characteristics to satisfy user the needs and achieve higher usability. This research examines the usability of common IR websites using two different usability evaluation methods; usability testing and system usability scale (SUS) technique

3 Research Methodology

The essence of usability evaluation is to find out what works and what does not work in an interface. There are different usability evaluation methods. Certain methods use data from users, while others rely on usability experts. In a recent study, Hass [51] had discussed several usability methods and their effectiveness. He described the steps involved in running a usability study such as users' recruitment and test environment. This can help to formulate how and what data to be collected. Majed and Mayhew [52] found that task design can play a vital role in usability testing results, where it was shown that changing the design of the task can cause differences in the results. Roobaea et al. [53] developed a new evaluation method using heuristic evaluation and user testing techniques in novel ways. Their adaptive methodological framework is proposed to be utilized for generating a domain-specific evaluation method. This has been claimed to offer better usability evaluation results in any certain domain [53].

3.1 Evaluation Methods

To evaluate the usability of common IR websites, we used two usability evaluation methods as recommended by Nielson Norman Group [54] and digital.gov [55], and commonly used by researchers [11,48,56]: (a) Usability Testing and (b) System Usability Scale (SUS).

3.1.1 Usability Testing

Usability testing is expected to offer a deep understanding for a certain website. It can also clearly help to better understand the usage and the user experience of user interfaces through experimental approach. Typically, during an experiment, participants perform a set of tasks while observers watch, listen, and take notes. The goals of the evaluation method are to identify any usability problems, collect data, and determine the participant's satisfaction with the product.

The usability testing technique we use is unmoderated and is conducted in-person. An unmoderated test is done without direct supervision. In this test, the researchers watch but are not allowed to participate, thereby minimizing the chance that a moderator could affect the users with their concerns. In-person tests offer additional data points, since observers are able to watch and analyze users' body language and their facial expressions.

The data points we gathered are quantitative in nature (i.e., frequency and severity rating of the problems faced, success rates, and task time). These types of data yield the participant's perceived objective difficulty and are popular for comparative study and, thus, are a best match for our research.

To complement the user's perceived objective difficulty, we also measure their user's subjective difficulty by using the Concurrent Think-Aloud (CTA) technique. The CTA is one of the dominant approaches of usability testing and it provides insight into people's experiences as they interact with the website [57,57]. During their task execution, participants are asked to verbalize what they are thinking and doing as they performed a task.

Further, the participant's testing session is timed and the audio and video of the task execution is recorded. Audio recorded the participants voices while verbalizing their CTA and video recorded the screen to capture participant's interaction with website, such as mouse clicks, movements, and scrolling. Session recordings are described as a key source to detect major usability problems and errors. It can also enrich the data as observers can determine how users interact with the page elements while performing the given tasks.

3.1.2 System Usability Scale (SUS)

After completing tasks, each participant evaluated the difficulty level for fulfilling the tasks (perceived difficulty a.k.a. subjective difficulty) by responding to System Usability Scale (SUS) form. The SUS is considered as an alternative and standardized subjective method for assessing the usability of a system from participants point of view. Participants respond to ten statements using a 5-point Likert scale (strongly agree to strongly disagree) [58]. A final score from 0 to 100 is calculated for the usability of the system. As interpreted by [59], the SUS score below 30 means that the website is not acceptable, a score between 30 and 60 means the website is marginally acceptable and above a score above 60 makes the website acceptable.

3.2 Selection of IR websites

To include all of the IR websites for the detailed analysis, it would be a tedious task. To obtain a broader picture of the qualities of IR websites, we classified the websites based upon their nature and feature and selected one popular representative IR from that category. The selected websites for the usability testing are; (a) ACM (www.acm.org/dl) – a publisher's digital library, (b) The Internet Archive (TIA) (www.archive.org) – an open access digital library, (c) Research Gate (RG) (www.researchgate.net) – a researcher community social networking website, (d) Semantic Scholar (SS) (www.semanticscholar.org) – an AI based search system, and (e) Google Scholar (GS) (scholar.google.com) – an open access search website. Figure 1 shows the landing page of these IR websites. See Table 1 for the list of the selected IR web systems and their characteristics.

(a) ACM

(b) TIA

(c) RG

(d) SS

(e) GS

Figure 1. The landing page across 5 IR Website

Articles produced by the Association for Computing Machinery (ACM) in its journals, magazines, and sessions are all available in full-text form in ACM. The Internet Archive (TIA) is a virtual library that makes available online for no cost a wide variety of digitized works. It has millions of e-books, hundreds of software programs, and thousands of websites. RG is a paid social network, where scientists and academics may exchange and discuss published articles. It links over 20 million academics and provides access to over 135 million papers. Powered by AI, SS helps users sift through volumes of academic material. It summarizes academic publications using the state-of-the-art NLP technology. GS is a publicly available online search engine that indexes the full text or metadata of academic literature from a wide variety of publishers and fields of study. These five IR sites were selected not only for their widespread use, but also for evaluates the various styles of information presentation, search, and retrieval.

ACM website presents the content in an asymmetrical layout whereas TIA uses a long scrolling grid of cards layout for the same. RG presents the content in a long scrolling single column layout. SS displays the contents in a layout similar to the F-shaped eye grazing pattern whereas GS uses a simple fixed sidebar web page layout.

Table 1. IR web systems included for the usability testing

Name	Category	Open Access?	Free/paid?	Coverage (million)
ACM	Publisher's digital library website	Proprietary	Paid	2
TIA	Open access digital library website	Open	Free	36
RG	Researcher's social networking website	Proprietary	Free	135
SS	AI based search website	Proprietary	Paid	40
GS	Open access search website	Mixed	Free	390

3.3 Selection of Testing Tasks

The way people interact with an IR website is highly contingent upon; (a) their task, (b) their assumptions from previous experiences, (c) the page layout, and (d) the type of the content. Having this in mind, the user's experience with an IR web site is highly subjective and unique. Nevertheless, there are necessary features that an IR website is expected to provide to fully expose its functionality. These necessary features are to enable a user to; (a) precisely specify a search query to easily find a piece of information, (b) sort the results of search based upon the desired characteristics, (c) download and fully explore a searched item, (d) easily find and use primitive website features, e.g., access to privacy policy, locate cost of subscription and make payment, and print a desired file or piece of information. To test the usability of these necessary features of the IR websites, we opted a user-centered task design approach similar to the [8]. This approach allows test participants to behave naturally and freely express their needs while performing the given tasks.

Table 2. Tasks and their respective Objectives and Explanation

Sr. No.	Task	Explanation
Task-1	Search a piece of information	Searching is the most primitive feature of any IR system and a well-designed IR website must support this feature at a prominent position.
Task-2	Download a piece of information	The searched results must be downloadable to be accessible for an offline reading.

Task-3	Print out piece of information	Besides the search capability, a well-designed IR website should also enable ease of use by providing primitive usability features such as printing.
Task-4	Find a price for a service	The paid services provided by the IR website must be clearly marked and accessible on the IR website.
Task-5	Explore a piece of information	The IR website must not depend upon external resources and should facilitate an enjoyable and satisfactory platform for the reading of the retrieved content.

We designed five tasks which are explained in Table 2. The rationale behind opting five tasks is to ensure that HCI usability recommendations [1,6] are met which suggests that the number of tasks should be between three and seven. In order to perform any job, users started at the homepage and then navigated to other sections of the site. Each task is designed in way that it requires exploration of a certain section on the website. Collectively these five tasks require participants to fully explore a website under a test. These tasks, therefore, enable us to better compare the usability pros and cons of the IR websites.

In this way, if one step in the process fails, it won't affect the rest of the process, and vice versa. Moreover, the activities were ordered such that each successive one would be more difficult to complete than the one before it. Each successive job will be completed more quickly because users are supposed to have learned from their prior experience, mitigating the effect of the increasing complexity of the tasks. In addition, a brief narrative has been provided to clarify the responsibilities and set the setting.

As all the IR websites under study, differ in their approach and features, one task cannot fit for all of them. For example, ACM requires a subscription to give a full access to its content, whereas the Internet Archive seeks a voluntary donation. We, therefore, slightly modified the tasks to adjust to the working of a given IR website. These tasks are described in Table 3.

Table 3. IR Websites and their respective Tasks

Website	Task	Description
ACM	**Task-1**	Find e-book *A practical guide to usability testing* by Joseph Dumas.
	Task-2	Find and download an article cited by more than 50 times.
	Task-3	Print the *Privacy Policy* statement of the website.
	Task-4	What’s the yearly library subscription cost for university students from China?
	Task-5	Find an article defining *contextual inquiries*.
TIA	**Task-1**	Find e-book *A practical guide to usability testing* by Joseph Dumas.
	Task-2	Find and download an article previewed more than 50 times.
	Task-3	Print the *Privacy Policy* statement of the website.
	Task-4	What is the minimum monthly donation one can make?
	Task-5	Find an article defining *contextual inquiries*.
RG	**Task-1**	Find e-book *A practical guide to usability testing* by Joseph Dumas.
	Task-2	Find and download an article cited by more than 50 times.
	Task-3	Print the *Privacy Policy* statement of the website.
	Task-4	How much it cost to post a 30-day job advertisement to hire two persons?
	Task-5	Find an article defining *contextual inquiries*.
SS	**Task-1**	Find e-book *A practical guide to usability testing* by Joseph Dumas.

	Task-2	Find and download an article cited by more than 50 times.
	Task-3	Print the *Privacy Policy* statement of the website.
	Task-4	What's costs for the RESTful S2AG API to look up Paper and Author?
	Task-5	Find an article defining *contextual inquiries*.
GS	**Task-1**	Find e-book *A practical guide to usability testing* by Joseph Dumas.
	Task-2	Find and download an article cited by more than 50 times.
	Task-3	Print the *Privacy Policy* statement of the website.
	Task-4	What's the cost to link your website with Google Scholar?
	Task-5	Find an article defining *contextual inquiries*.

3.4 Participants

Usability testing guidelines [54,55] suggest that quantitative studies should test approximately 15 to 20 users to get statistically significant numbers. We, therefore, recruited 18 voluntary users – ten students, five faculty members, and three staff members. There were ten males and eight female participants. Each participant performed these five tasks on all five IR websites. The participants and their characteristics are presented in Figure 2.

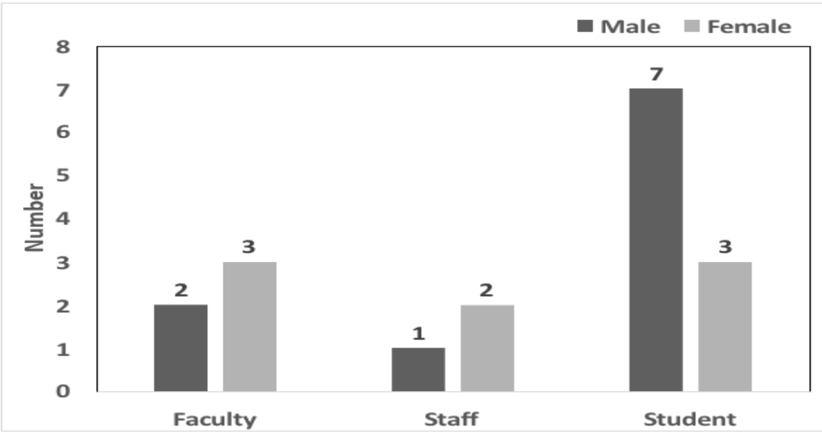


Figure 2. The Participants and their Characteristics

3.5 Metrics and Measures

During the task planning, candidate selection, task execution, and after completion of the testing sessions via analysis of observations and verbalization recorded by CTA, we accumulate usability testing quantitative data. The analysis of these quantitative usability testing data helps us to describe the user experience and identify the design elements and structures of information that formulate a good design increases the usability of an IR website. The metrics we aim to measure, are summarized in Table 4 and explained in details in the rest of this section.

Table 4. Usability Testing Measures

Metric	Explanation
Problem	Number of common usability problems
Frequency	Severity Rating of each problem Number users encountered each problem
Success Rates	**Success Rate at Level-1** Percentage of users who completed a task without any help **Success Rate at Level-0.5** Percentage of users who completed a task with a help **Success Rate at Level-0** Percentage of users who failed to complete the task Task with highest success rate at each level Task with lowest success rate at each level
Task Times	**Average Task Success-Time** Report time spent by users who successfully completed the task **Average Task Failure-Time** Include time spent by users who failed to complete the task **Average Task Time** Measures the total duration users spent on the task

3.5.1 Problem Frequencies

It's important to know what are the common usability problems, what are their severity ratings, and how many users encountered those problems, therefore, we computed the usability problem frequency [60]. Firstly, to identify the unique problems faced by participants, we analyzed the testing sessions and CTA verbalization. The approach we used to analyze the testing session is proposed by McCloskey [61] and is widely practiced in the industry [62–64]. These problems are also known as measured difficulty or objective difficulty. As not all the problems are the same and have different effects on the usability of the website, we also assign a severity rating to every problem. The severity of the usability problems is defined by Nielsen [64] as a combination of three factors; (a) the frequency with which the problem occurs, (b) the impact of the problem if it occurs, and (c) the persistence of the problem. Based on their severity ratings, these usability problems can be classified into five different categories (0 for the least important, and 4 for the most important), as shown in Table 5. Lastly, we count how many users encountered each of these unique usability problems. This measure indicates how common a usability problem is.

Table 5. Problem Severity Rating

Rating	Category	Description
0	No Problem	This is not a usability problem at all
1	Cosmetic	This only needs to be fixed if time permits
2	Minor	Fixing this should be given low priority
3	Major	Fixing this should be given high priority
4	Fatal	Must be fixed

3.5.2 Success Rates

One of the more common metrics used in a user experience is a task success or completion rate. We used a three levels of success [65] measure: (a) Success Level-1: when a user was able to complete a task without any help, (b) Success Level-0.5: when user was able to complete a task with support or with lots of effort, and (c) Success Level-0: when a user failed to complete a task. To report the success rate, we calculated the percentage of users who were at a given success level. For each success level, we then reported the tasks with the highest and lowest success rates. The task with the highest success rate depicted the most convenient task and the task with lowest success rate indicated the most cumbersome task.

3.5.3 Task Times

Time spent on a task is the most commonly collected metric for usability studies. A task time measures the efficiency and productivity of a product or a service [66]. If the user spends an extended amount of time completing an action, that means the interaction is not properly designed. We measure the time as the seconds a participant spend to successfully complete a given task with or without an assistance. If the user fails to complete the task, we measure the task time as the time a user spent on the task. We measured following three task time measures [67]: (a) Average Task Success Time which includes only users who have successfully completed the task, (b) Average Time to Failure which reports the average time that the users are spending on a task before they give up or complete the task incorrectly, and (c) Average Time on Task which measures the total duration users are spending on a task.

4. Usability Tests Findings and Analysis

As mentioned earlier, this study sought to answer the research questions as set out in Section 2.1. In this section, we evaluate and analyze the testing results to be able to examine the IR website's usability.

4.1. Usability Testing Results

4.1.1. Frequencies of Usability Problems

It has been identified the unique problems faced by the participants and then calculated the frequency of the problems while performing the given tasks. Experts analyzed the usability audio and video sessions and identified a total of five unique problems. Of the five unique problems discovered, P-1 problem was rated as not a usability problem of Severity-0, P-2 as a cosmetic problem of Severity-1, P-3 and P-4 as major problems of Severity-3, and P-5 as

fatal problem of Severity-5 – no minor rated problem of Severity-2 was found during our analysis. We wanted to compute frequencies of the Usability Problem. This measure indicates the difficulty level of participants while performing a given task. Firstly, for each unique problem, we counted its occurrences in an execution of a given task. Then, we computed the problem occurrence to a participant count ratio by dividing these occurrences with the number of participants encountered that problem in that given task. These Frequencies of Usability Problem along with their severity ratings are shown in Table 6.

Table 6. Frequencies of Usability Problem Faced

Sr. No.	Problem Observed	Severity	ACM	TIA	RG	SS	GS
P-1	Task performed without an issue	Severity-0	2	5	36	56	49
P-2	UI could have been improved	Severity-1	79	76	40	27	39
P-3	It took too many clicks	Severity-3	22	2	7	1	0
P-4	It was not easy to locate the feature	Severity-3	52	55	30	6	24
P-5	Did not complete the task	Severity-4	42	41	20	10	22

The results show that participants observed most the problems while interacting with ACM website and the minimum problems on SS website while performing the five tasks. According to these results, participants observed most Severity-1, Severity-3, and Severity-4 problems while interacting with ACM website and minimum Severity-1, Severity-3, and Severity-4 problems while performing tasks on SS website.

The frequency of problem severity was also calculated for each task. This measure indicates how catastrophic usability experience of participants is while performing a given task and the results are presented in Table 7.

As seen in Table 7 that while performing the Task-1 on ACM website, 17 participants observed at least one or more Severity-1 problem and all the 18 participants faced at least one or more Severity-3 and Severity-4 problems. In comparison, on SS website, 3 participants observed at least one or more Severity-1 problems and two participants faced at least one or more Severity-4 problems. The results show that participants found it very convenient to perform Task-1 on SS website and most difficult on ACM website. As Task-1 tests the search capability of the IR websites, we can assert that SS presents the search results in the most usable and convenient manner.

The results suggest that participant found it most difficult to perform Task-2 on TIA website (18 participants faced Severity-1 problems, two observed Severity-3 issues and eight faced Severity-4 problems) and easiest on the SS website (three participants were observed to face Severity-1 problems only). It shows that the SS website also delivers the best experience while performing the Task-2 of downloading a particular piece of information.

While participants were performing the Task-3, experts observed the results similar to Task-2. While performing Task-3 on TIA website, 15 participants observed at least one or more Severity-1 problem, 14 participants faced at least one or more Severity-3, and 11 participants were seen facing Severity-4 problems. In comparison, on the SS website five participants observed at least one or more Severity-1 problem and only one participant faced at least one or more Severity-3 and Severity-4 problems. Again, SS website the outperformed other IR websites when participants performed the task of printing.

The participants found RG website most cumbersome to deal with while performing Task-4 (17 participants observed at least one or more Severity-1 problem, all 18 participants faced at least one or more Severity-3, and 10 participants were seen facing Severity-4 problems) and is SS easiest (nine participants observed at least one or more Severity-1 problem, two participants faced at least one or more Severity-3, and four participants were seen facing

Severity-4 problems). The results show that the participants experience the highest usability while searching the price of a piece of information.

Table 7. Frequency of Problem Severity and their Aggregate

Severity	Task	ACM	TIA	RG	SS	GS
Severity-0	Task-1	0	0	13	14	9
	Task-2	2	0	5	15	13
	Task-3	0	0	7	13	11
	Task-4	0	5	0	3	0
	Task-5	0	0	11	11	16
Severity-1	Task-1	17	18	3	3	8
	Task-2	12	18	7	3	5
	Task-3	18	15	7	5	7
	Task-4	14	11	17	9	15
	Task-5	18	14	6	7	4
Severity-2	Task-1	0	0	0	0	0
	Task-2	0	0	0	0	0
	Task-3	0	0	0	0	0
	Task-4	0	0	0	0	0
	Task-5	0	0	0	0	0
Severity-3	Task-1	27	16	2	0	5
	Task-2	11	14	3	0	0
	Task-3	6	14	3	1	5
	Task-4	20	6	24	3	14
	Task-5	13	6	5	0	0
Seveirty-4	Task-1	18	7	1	2	2
	Task-2	18	8	4	0	1
	Task-3	2	11	3	1	1
	Task-4	5	7	10	4	16
	Task-5	9	8	2	3	2

Finally, the ACM website was found to be the most difficult to interact with while performing the Task-5 (all 18 participants observed at least one or more Severity-1 problem, 13 participants faced at least one or more Severity-3, and nine participants were seen facing Severity-4 problems) and the GS is the most convenient one (four participants observed at least one or more Severity-1 problem, and two participants were seen facing Severity-4 problems). The participants have almost similar experiences at the SS website where six participants observed at least one or more Severity-1 problem, and three participants were seen facing Severity-4 problems.

Table 7 also shows the aggregate of these frequencies. The ACM website was found to be the most poorly structured and the participants found a highest number of Severity-1, Severity-3 and Severity-4 problems with this website, a total of 79, 77, and 42, respectively. The SS website performed the best among all these websites with a total of 27 Severity-1, four Severity-3 and a total of ten Severity-4 problems. The results show that the ACM website is most prone to the problems while SS is the least prone to them.

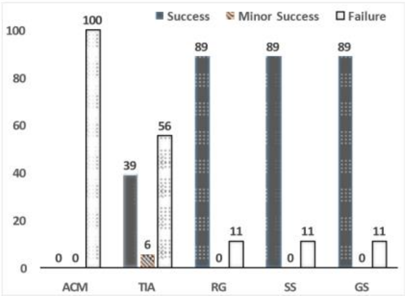
These findings revealed that SS outperformed all other IR websites in terms of the overall number of usability problems discovered, and the number of unique and minor problems for each task performed. An explanation for these results may be that the SS website not only yielded better results to the search query but also presented them in a very convenient and easy way to locate layout. Not surprisingly, GS and the RG websites having similar landing pages performed close to the SS website and better than TIA and ACM websites.

4.1.2 Success Rates of Tasks

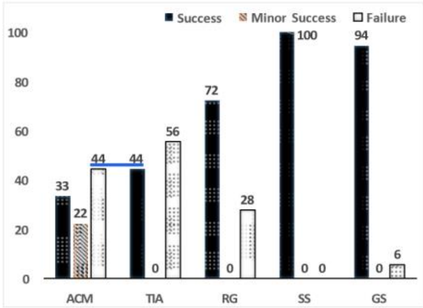
Firstly, we calculated the Success Rates at different levels of success for each task. The results are shown in Table 8 and Figure 3.

Table 8: Success Rate for each Task across all IR Websites

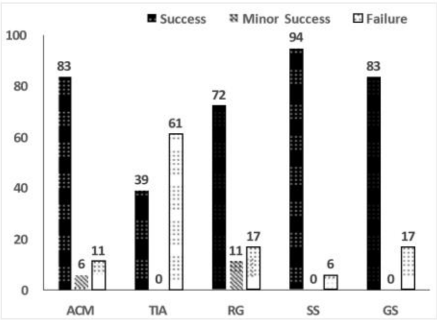
Task	**ACM**			**TIA**			**RG**			**SS**			**GS**		
	S	**M**	**F**	**S**	**M**	**F**	**S**	**M**	**F**	**S**	**M**	**F**	**S**	**M**	**F**
Task-1	0	0	0	39	6	56	89	0	11	89	0	11	89	0	11
Task-2	33	22	44	44	0	56	72	0	28	100	0	0	94	0	6
Task-3	83	6	11	39	0	61	72	11	17	94	0	6	83	0	7
Task-4	56	11	33	61	0	39	11	28	61	78	0	22	11	0	89
Task-5	0	0	100	61	0	39	83	0	17	83	0	17	94	0	6



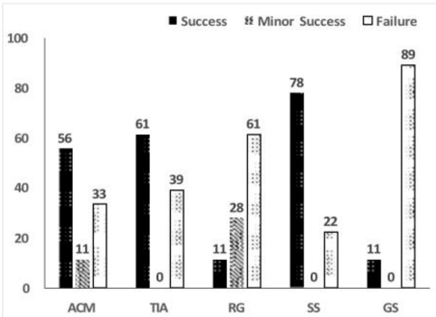
(a) Task-1



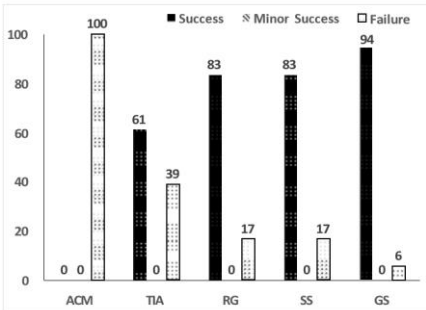
(b) Task-2



(c) Task-3



(d) Task-4



(e) Task-5

Figure 3. Success Rate for each Task across all IR Websites

The results show that the none of the participants was able to successfully complete Task-1 on the ACM website. The highest percentage of participants, who were (89%) successfully completed the Task-1 on the RG, SS, and GS websites.

The lowest percentage of participants (33%), successfully completed the Task-2 on ACM website whereas all the participants successfully completed this task on SS website. The 39% of the participants struggled to successfully complete the Task-3 on TIA website. On the other hand, 94% of participants completed the same task on the SS website without any help.

Task-4 was found to be the most difficult task where 56%, 61%, 11%, 78%, and 11% of the participants successfully completed this task on the ACM, TIA, RG, SS, and GS websites, respectively. Also, for this task the maximum number of the participants (28% for RG website) demanded assistance in comparison with any other task.

Finally, 94%of the participants were successfully completed the Task-5 on GS website and 83% of the participants successfully completed this task on the RG and SS websites. No participant was able to complete this task on the ACM website.

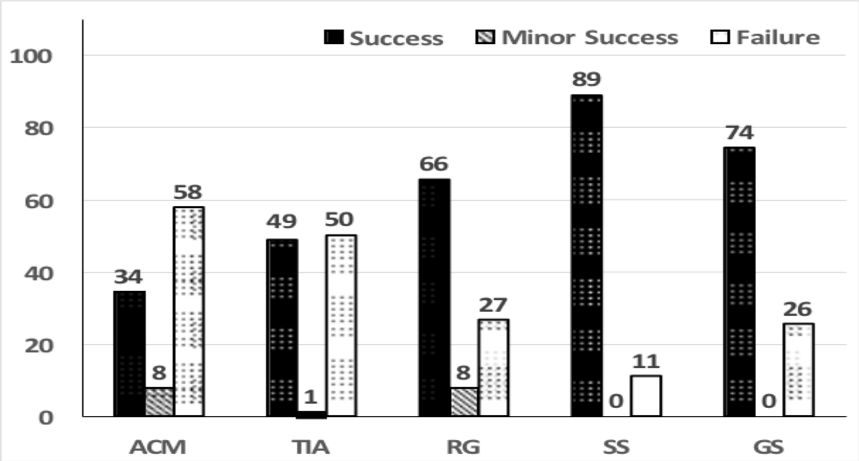


Figure 4. Success Rate for each IR Website

The results in Figure 4 show the average success rate of all five tasks for each IR website. ACM website yielded the poorest usability success rates. On the ACM website for all of the five tasks, the average level-1 success rate is 34% and the average level-0.5 success rate is 8%. On the other hand, 58% of the participants failed to complete the given tasks on the ACM website. Level-1 and level-0.5 success rates for all the given tasks performed on the TIA website are 49% and 1%, respectively whereas for the same the level-0 success rate is 50%.

For all the given tasks performed on the RG website, the Level-1 and level-0.5 success rates are 66% and 8%, respectively whereas the level-0 success rate for all the tasks is 27%. The participants most successfully performed their tasks on the SS website. For this website Level-1 and level-0.5 success rates are 89% and 0%, respectively. Only 11% of the participants failed to perform the given tasks on the SS website.

GS website delivered the second-best level-1 success rates (74%) and 0% level-0.5 success rates. On this website, 26%of the participants failed to perform the given tasks.

4.1.3 Time on Tasks

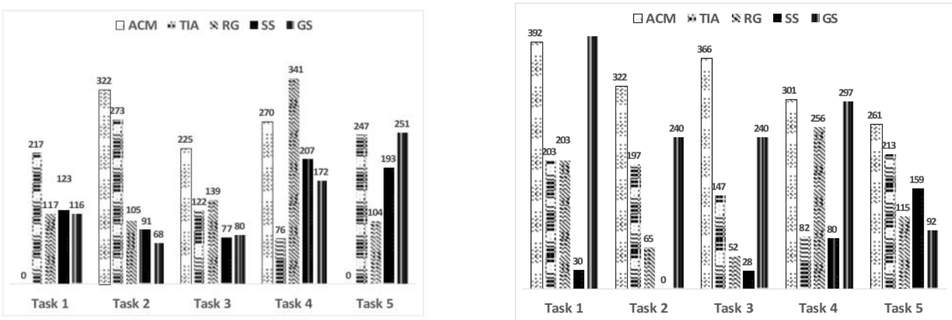
The audio and video recordings of the testing session have been analyzed and the time spent on the tasks has been calculated. The results are shown in Table 9.

Table 9: Average Success(S) and Failure(F) Time**(sec)** of Tasks across 5 IR Website

IR Websites	**Task-1**		**Task2**		**Task--3**		**Task-4**		**Task-5**	
	S	**F**	**S**	**F**	**S**	**F**	**S**	**F**	**S**	**F**
ACM	0	392	322	322	225	366	270	301	0	261
TIA	217	203	273	197	122	147	76	82	247	213
RG	117	203	105	65	139	52	341	256	104	115
SS	123	65	91	0	77	28	207	80	193	159
GS	116	0	68	240	80	240	172	297	251	92

Firstly, we compute the Average Task Success-Time which includes only users who have successfully completed the task. The results are presented graphically in Figure 5(a). The minimum average success time suggests how convenient a task was for the participants to complete. The minimum average success time to complete Task-1 is 116 seconds and is observed on the GS website. That is followed by RG and SS websites with average success time 117 seconds and 123 seconds, respectively. The minimum average success time for Task-2 is 68 seconds on GS followed by SS and RG with average success time 91 seconds and 105 seconds, respectively. The minimum success time for Task-3 is 77 seconds on SS, for Task-4 is 76 seconds on TIA, and for Task-5 is 104 seconds on RG. No participant could complete the Task-1 and Task-5 on the ACM website.

Secondly, we measured the Average Task Failure-Time which is also known as Mean Time to Failure. It reports the average time users are spending on the task before they give up or complete the task incorrectly. The Figure 5(b) presented the results. The minimum average task failure time for Task-1 is 30 seconds and is observed on the SS website. The minimum average task failure time for Task-2 is 65 seconds on RG. No participant failed to perform Task-2 on the SS website, and therefore, the SS website does not have an Average Task Failure-Time. The average task failure time for Task-3 is 28 seconds on SS, for Task-4 is 80 seconds on SS, and for Task-5 is 92 seconds on GS.



(a) Average Success Time

(b) Average Failure Time

Figure 5. Average Time of Tasks across 5 IR Websites

We also report the Average Time on Task in Figure 6. The minimum average task time for Task-1 is 112.5 seconds and is observed on SS website. The same time for Task-2 is 77.4 on GS, for Task-3 is 74.39 on SS, for Task-4 is 78.01 on TIA, and for Task-5 is 106 on SS.

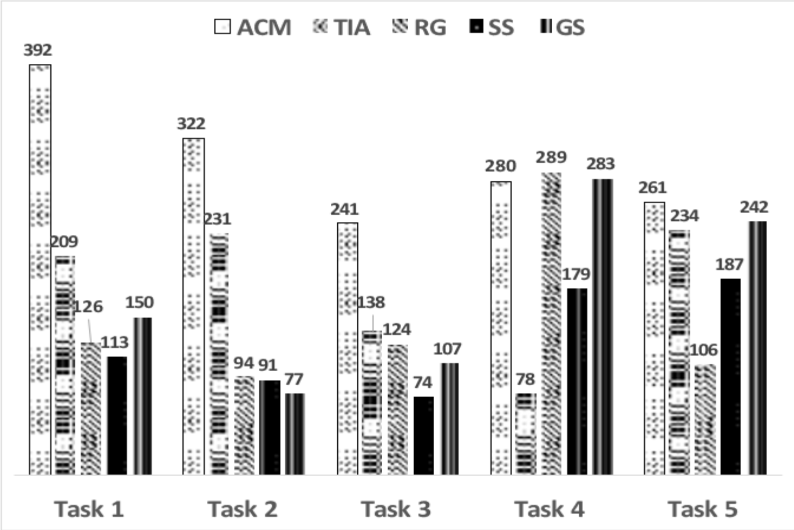


Figure 6. Average Task Time for IR Websites

There is a large amount of variation in the time to complete the task on an IR website. Such behavior patterns may have many origins: previous experience in use of the particular service, knowledge of the similar IR web systems, or the knowledge of the domain in which the tasks were framed. These results are in line with a number of other studies such as [41,68,69], which suggest that the use of familiar information alongside location of the elements of a web design can have a positive impact of the visual search, thus increasing the chances for a faster task performance.

4.2 SUS Results

In order to measure participants' perceptions of the usability of the chosen website, participants were asked to fill out the SUS. An analysis of the responses from both groups is summarized in Figure 7.

The overall analysis reveals that participants found the ACM website very unusable, unnecessarily complex and cumbersome giving it merely 13 on the SUS scale. The three websites (i.e., RG, SS and GS) all received high scores in the proximity of 90 on the SUS scale. The SUS results slightly differ from the problem severity results presented in Table 6 and in Table 7. That is due to the fact that participants rate problems are different from the experts. The experts have knowledge of the problem frequency, problem statements, and better understand their severity levels on the overall user interface.

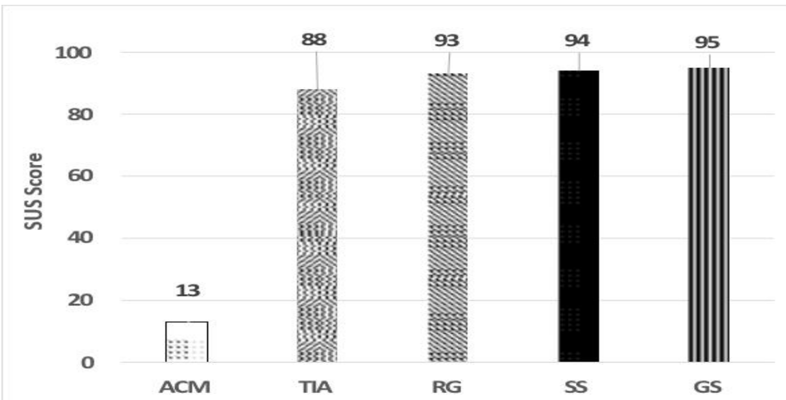


Figure 7. SUS Results across 5 IR Website

Compared to earlier research it has been found that for the measurement of the usability of websites and applications, SUS is widely adopted due to its remarkable advantages -it provides reliable results and is easy to conduct, as compared to the other surveys [70-73]. Many users reported that it was difficult to navigate and find features on ACM, making the system unnecessarily complex. This issue has also been acknowledged by earlier research that indicates that the most frustrating challenge screen reader users face on the web is navigation [74]. Other prior works [75-77] also point out that the issues that make systems unnecessarily complex, such as navigation issues, should be considered in the web educational systems.

5. Discussion and general findings

This section explores the analysis and discussion of the usefulness and ease of use of the user interface elements and their placement on the IR web page.

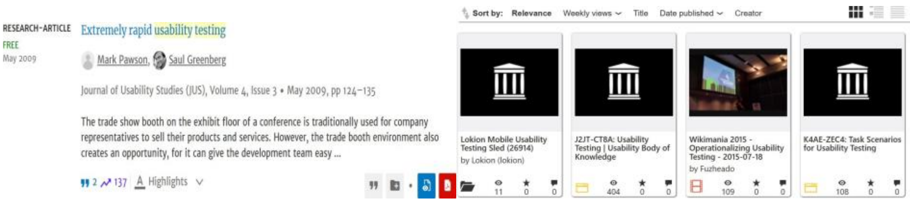
Task-1 was designed to test the usability of the most primitive feature of an IR website: searching. Moreover, it is generally anticipated that the search feature on an IR website should be at a highly visible section of the page. So, our motivation was also to find out how the past experience of a participant affects the usability testing of the IR website. This research finding emphasizes that even if a participant's past experience affects their experience on a website, it has been proposed that the design of the webpage itself has a greater influence on the webpage usability.

Of all the IR websites we evaluated, the websites with the full-screen search bar (i.e., RG, SS and GS) yielded the best success rates (89% each for RG, SS, and GS) and the best average task success time (116 seconds for GS, 117 seconds for RG, and 123 seconds for SS) for Task-1.

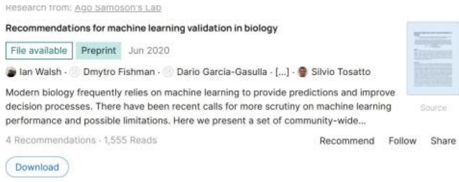
On the other hand, the ACM website provides two search capabilities: searching a feature on its website and searching the content in its digital library. The search bar needed to search in the digital library is embedded inside its website many clicks away, whereas the search bar to find features on its website is provided on its main landing page. Almost all of the participants who wanted to search the required content in the ACM digital library were misled and ended up on the wrong search bar. The few participants who found the correct search bar were still not able to search the required content in the digital library due to the poor searching capability of ACM search engine. Consequently, none of the participants were successful in the Task-1.

Moreover, TIA website has multiple search bars on its main page. Two search bars are not as bad as it might sound, as long as they are clearly separate and labeled. However, TIA failed to do so and, therefore, achieved only 39% success rate for Task-1.

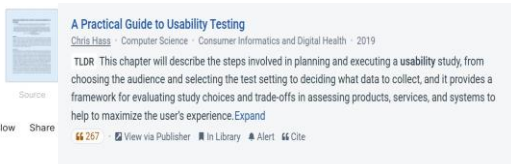
Task-2 was designed to test the sorting, selection, and accessibility of the searched results. We asked participants to download certain information retrieved against a search query. The output screen for all five IR Websites against Task-2 is shown in Figure 8.



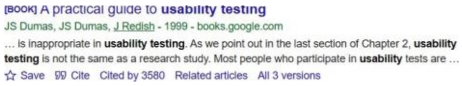
(a) ACM (b) TIA



(c) RG



(d) SS



(e) GS

Figure 8. Search result items for Task-2 across Five IR Websites

Our finding revealed that all the websites (i.e., SS and GS) which provided an upfront and easily accessible citation count of the searched results achieved high success rates (100% for SS, and 94% for GS).

RG provided an option to sort the retrieved results based upon the citation count that was not easily viewable and, therefore, received a comparatively a lesser success rate of 72%. On the other hand, the the ACM website also provided the citation counts readily available as did SS and the GS. It, however, failed to receive a high success rate for Task-2 as well the due to the difficulty in locating the search bar in the first place which hindered progress.

Similarly, the the TIA website provided readily accessible information about the number of times that the searched item has been previewed. However, TIA also only received 44% of success rate as its search engine failed to retrieve the desired information.

In Task-3 which was designed to test the usability of the IR website for providing the print feature, participants were asked to access the privacy policy of the IR website and print it. Moreover, it is expected that the privacy policies are generally placed at the bottom of the page. Therefore, as was the case while designing the Task-1, our motivation was also to find out how the past experience of a participant impacted the usability testing. The usability testing results of Task-3 also confirmed that the design of the webpage has a greater influence on webpage usability as compared to past experience. ACM, SS, and GS provided the privacy policy on the bottom of the page where it is generally expected and, therefore, these websites received high success rates, 83%, 94%, and 83%, respectively. Whereas, the TIA and RG website presented its content in a continuous scrolling layout which prevented participants from reaching the bottom of the page. These two websites, therefore, received the lowest success rate among the five IR websites we tested.

Participants have been asked to find the price of the service offered by the website in Task-4. Most of the IR websites performed poorly against this task. This is due to the fact that the majority of websites keep the price tag hidden to avoid unfair competition from rivals and to control the tendency of driving prices down [78-81]. Therefore, most of the participants were not easily able to find the price of the service and failed.

Finally, Task-5 was designed to test the IR website for their ease of use to fully explore a searched piece of information. The usability results of Task-5 show that GS, SS, and RG websites give the best usability experience with a success rate of 94%, 83%, and 83%, respectively. It is partially due to their better search engine which retrieved the accurate results.

The TIA website received a success rate of 61% only mainly because of its grid of card page layout. It is in contrast to the norm where participants expect the results to be displayed in a single column layout. Again, participants failed to perform Task-5 on the ACM website due to their inability to find the ACM Digital Library search page and its poor search engine which failed to accurately find the desired information.

6. Conclusion

These research results assert that users may find an accurate IR system difficult to circumvent if it is not usable. We analyze and discussed the usefulness and ease of use of the user interface elements and their placement on the IR web page. Even if a user's past experience influences their experience on a website, it is confirmed that the design of the webpage itself has a greater impact on the usability of webpage.

These research results suggest that the location of the feature on the webpage plays an important role as far as the efficiency of the task is concerned. For example, the SS website yields quickest time for the searching task (i.e., Task-1) due to placement of the search bar on the most convenient position on the webpage. Similarly, GS outperforms other websites for the downloading task (i.e., Task-2) because it allows users to directly download the PDF file with a single click next to the searched item.

Our findings also confirmed and revealed that failing at a number of necessary conditions does not mean that the search system should be avoided entirely in the systematic review process, but can imply that it should not be used for query-based searching. Nevertheless, such websites might be used in supplementary search methods. The SUS evaluation score of 79, which corresponds to "acceptable", whereas the initial score was 66. The average score for SUS, remote usability testing after the second-round evaluation is increased. We considered our results valuable for web design development.

Funding: This research was funded by the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia, grant number 523.

References

1. Norm, D. Ergonomics of Human-System Interaction. Part II: Usability: Definitions and Concepts. Standard, International Organization for Standardization, 2018. Available at: https://www.iso.org/standard/63500.html (Accessed: March 28, 2023).
2. Dubey, S.K.; Rana, A. Analytical Roadmap to Usability Definitions and Decompositions. IJSER 2010, 2, 4723 – 4729.
3. Palmer, J. Web Site Usability, Design, and Performance Metrics. ISR 2002, 13, 151–167.
4. Younghwa, L.; Kenneth A., K. Understanding of Website Usability: Specifying and Measuring Constructs and Their Relationships. DSS 2012, 52, 450–463.
5. Bevan, N. International Standards for HCI and Usability. INT J HUM-COMPUT INT 2001, 55, 533–552.
6. Jacob, N. Usability Engineering; Academic Press Boston, 1993.
7. Jeng, J. What Is Usability in the Context of the Digital Library and How Can It Be Measured. LITA 2005, 24, 3–12.
8. Tarkkanen, K.; Reijonen, P.; Tetard, F.; Harkke, V., Back to User-Centered Usability Testing. LNCS, 2013; Vol. 7946, pp. 91–106.
9. Wang, Q.; Yang, S.; Liu, M.; Cao, Z.; Ma, Q. An Eye-Tracking Study of Website Complexity from Cognitive Load Perspective. DSS 2014, 62, 1–10.
10. Sonderegger, A.; Sauer, J. The Influence of Design Aesthetics in Usability Testing: Effects on User Performance and Perceived Usability. Appl. Ergon. 2010, 41, 403–410.
11. Tichindelean, M.; Tichindelean, M.; Cetina, I.; Orzan, G. A Comparative Eye Tracking Study of Usability - Towards Sustainable Web Design. SUSTDE 2021, 13, 10415.
12. Labrecque, L.I.; Milne, G.R. Exciting Red and Competent Blue: The Importance of Color in Marketing. JAMS 2012, 40, 711–727.
13. Gorn, G.; Chattopadhyay, A.; Sengupta, J.; Tripathi, S. Waiting for the Web: How Screen Color Affects Time Perception. J. Mark. 2004, 41, 215–225.

14. Ren, L.; Chen, Y. Influence of Color Perception on Consumer Behavior. In Proceedings of the HCI in Business, Government, and Organizations. Springer International Publishing, 2018, pp. 413–421.
15. Arabi, A. Influence of Colors on Consumer Behavior, "Conceptual and Theoretical Approaches". Annals of Constantin Brancusi University of Targu-Jiu. Economy Series 2017, 1, 163–170.
16. Nielsen, J. Designing Web Usability. J. 2000.
17. Mothersbaugh, L.D.; Hawkins, I.D. Consumer Behavior: Building Marketing Strategy; McGraw Hill, 2016.
18. Rory, H.; Bhasker, M. The Role of Atmospherics in Influencing Consumer Behaviour in the Online Environment. IJBSSR 2011, 2.
19. Bonnardel, N.; Piolat, A.; Bigot, L.L. The Impact of Colour on Website Appeal and Users Cognitive Processes. Displays 2011, 32, 69–80.
20. Laeng, B.; Suegami, T.; Aminihajibashi, S. Wine labels: An Eye-Tracking and Pupillometry Study. Int. J. Wine Bus. Res. 2016.
21. Rik, P.; Michel, W. Attention Capture and Transfer in Advertising: Brand, Pictorial, and Text-Size Effects. J. Mark. 2004, 68, 36–50.
22. Robert.; Passagno. Design and Usability of Social Networking Web Sites. Graphic Communication 2010.
23. Matzen, L.E.; Haass, M.; Divis, K.; Stites, M. Patterns of Attention: How Data Visualizations are Read. In Proceedings of the Augmented Cognition. Neurocognition and Machine Learning: 11th International Conference, AC 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings, Part I 11. Springer, 2017, pp. 176–191.
24. Zlokazova, T.; Blinnikova, I.; Grigorovich, S.; Burmistrov, I. Search Results on Flight Booking Websites: Displaying Departure and Return Flights on a Single Page vs Two Consecutive Pages. In Proceedings of the International Conference on Human Computer Interaction, 2019, pp. 668–671.
25. Gossen, T.; Nurnberger, A. Specifics of information retrieval for young users: A survey. Inf Process Manag 2013, 49, 739–756.
26. Zahabi, M.; Kaber, D.; Manida, S. Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literature and Guideline Formulation. Hum. Factors 2015, 57, 805–834.
27. Anupriya.; Kaur. Perceived Website Efficacy for Life Insurance Companies: Insights from a Best-Worst Method. IJITPM 2022, 13, 1–21.
28. Jiuqiang.; Fu. Usability Evaluation of Software Store Based on Eye-Tracking Technology. In Proceedings of the 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference. IEEE, 2016, pp. 1116–1119.
29. Bataineh, E.; Mourad, B.A.; Kammoun, F. Usability analysis on Dubai e-government portal using eye tracking methodology. In Proceedings of the Computing Conference. IEEE, 2017, pp. 591–600.
30. Kaysi, B.; Topaloglu, Y. Competitive Usability Testing of Student Information Systems with Eye Tracking Method. In Proceedings of the International Conference on Computational Science and Computational Intelligence, 2017, pp. 951–956.
31. Pawel, W.; Krzysztof, R.; Garnik, I. Eye-tracking web usability research. 2016 Federated Conference on Computer Science and Information Systems (FedCSIS) 2016, pp. 1681–1684.
32. Duygu, A.; Kursat, C. Analyzing Turkish E-Government Websites by Eye Tracking. In Proceedings of the J., 2013, pp. 225–230.
33. Sauer, J.; Sonderegger, A.; Heyden, K.; Biller, J.; Klotz, J.; Uebelbacher, A. Extra-Laboratorial Usability Tests: An Empirical Comparison of Remote and Classical Field Testing with Lab Testing. Appl. Ergon. 2019, 74, 85–96.
34. Diane.; Kelly. Methods for Evaluating Interactive Information Retrieval Systems with Users. Found. Trends Inf. Retr. 2009, 3, 1–224.
35. Savolainen, R.; Kari, J. User-defined Relevance Criteria in Web Searching. J Doc. 2006, 62, 685–707.
36. Kengeri, R.; Seals, C.; Harley, H.; Reddy, H.; Fox, E. Usability study of digital libraries: ACM, IEEE-CS, NCSTRL, NDLTD. Int. J. Digit. Libr. 1999, 2, 157–169.
37. Judy.; Jeng. Usability Assessment of Academic Digital Libraries: Effectiveness, Efficiency, Satisfaction, and Learnability. Libri 2005, 55.

38. Brenda, B.; Austin, B.; Weintrop, J. Usability testing of an academic library Web site: a case study. J. Acad. Librariansh. 2001, 27, 188–198.
39. Kelly L, M.; Diane H, S. User Perspectives on Relevance Criteria: A Comparison Among Relevant, Partially Relevant, and not-Relevant Judgments. JASIST 2002, 53, 327–342.
40. Carol.; Barry. User-Defined Relevance Criteria: An Exploratory Study. J. Am. Soc. Inf. Sci. 1994, 45, 149–159.
41. Casalo, L.; Flavian, C.; Guinaliu, M. The Role of Perceived Usability, Reputation, Satisfaction and Consumer Familiarity on the Website Loyalty Formation Process. Comput. Hum. Behav. 2008, 24, 325–345. Part Special Issue: Cognition and Exploratory Learning in Digital Age.
42. Khabsa, M.; Giles, C. The Number of Scholarly Documents on the Public Web. PloS One 2014, 9, e93949.
43. Thorpe, A.; Lukes, R. A Design Analysis of Indiana Public Library Homepages. PLQ 2015, 34, 134-161.
44. Anderson, T. Searching for Information: Applying Usability Testing Methods to a Study of Information Retrieval and Relevance Assessment. AARL 1999, 30, 189–199.
45. Kammerer, Y.; Gerjets, P. The role of thinking-aloud instructions and prior domain knowledge in information processing and source evaluation during Web search. In Proceedings of the J., 2013.
46. Manzari, L.; Trinidad-Christensen, J. User-Centered Design of a Web Site for Library and Information science Students: Heuristic Evaluation and Usability Testing. Inf. Technol. Libr. 2013, 25, 163–169.
47. Gusenbauer, M.; Haddaway, N. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. JRSM 2020, 11, 181–217.
48. Burnett, G.; Ditsikas, D. Personality as A-Criterion for Selecting Usability Testing Participants. J. 2006, pp. 599–6004.
49. Kortum, P.; Oswald, F.L. The Impact of Personality on the Subjective Assessment of Usability. INT J HUM-COMPUT INT 2018, 34, 177–186.
50. Bucher, H.J.; Schumacher, P. The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. J. 2006, 31, 347–368.
51. Hass, C., A Practical Guide to Usability Testing. In Consumer Informatics and Digital Health: Solutions for Health and Health Care; Springer International Publishing: Cham, 2019; pp. 107–124.
52. Majed, A.; Mayhew, P. Task Design: Its Impact on Usability Testing. In Proceedings of the J., 2008, pp. 583–589.
53. Roobaea, A.; Ali H, A.B.; P.J., M. Generating a Domain Specific Inspection Evaluation Method through an Adaptive Framework. International Journal of Human Computer Interaction (IJHCI) 2013, 4, 88.
54. Nielsen, J.; Norman, D. Nielsen Norman Group - Home. https://www.nngroup.com/, 1994.
55. Usability.gov - Home. https://www.usability.gov/, 1994.
56. Alnashri, A.; Alhadreti, O.; Mayhew, P. The Influence of Participant Personality in Usability Tests. IJHCI 2016.
57. Roobaea, A.; Ali H, A.B.; Mayhew, P. The Impact of the Combination between Task Designs and Think-Aloud Approaches on Website Evaluation. JSSD 2013, 2013, 1.
58. Brooke, J. SUS: A Quick and Dirty Usability Scale. Usability Evaluation in Industry 1996, 189, 4—7.
59. LeSuer, S. Measuring usability metrics. https://slickplan.com/blog/measuring-usability-metrics, 2022.
60. Sauro, J. How common are usability problems. https://measuringu.com/problem-frequency/, 2010.
61. McCloskey, M. Turn User Goals into Task Scenarios for Usability Testing. https://www.nngroup.com/articles/task-scenarios- usability-testing/, 2014.
62. Jacob, N. Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. In Proceedings of the Cost-justifying Usability. Citeseer, 1994, pp. 242–272.
63. Patrick W, J. An introduction to Usability; Crc Press, 1998.

64. Nielsen, J. Severity Ratings for Usability Problems. https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability- 709 problems/, 1994.
65. Nielsen, J.; Budiu, R. Success Rate: The Simplest Usability Metric. https://www.nngroup.com/articles/success-rate-the-simplest- usability-metric/, 2021.
66. Sauro, J. What To Do with Task Times When Users Fail a Task. https://measuringu.com/failed-times/, 2010.
67. Sergeev, A. User interfaces design and UX/Usability Evaluation. http://ui-designer.net/usability/efficiency.htm, 2010.
68. Reicher, G.; Snyder, C.; Richards, J. Familiarity of background characters in visual scanning. J. Exp. Psychol.: Hum. Percept. Perform. 1976, 2, 522.
69. Wang, Q.; Cavanagh, P.; Green, M. Familiarity and pop-out in visual search. Percept. psychophys. 1994, 56, 495–500.
70. Hamid, S.; Bawany, N. Z.; Zahoor, K. "Assessing ecommerce websites: Usability and accessibility study," 2020 International Conference on Advanced Computer Science and Information Systems, ICACSIS 2020, Oct. 2020, pp. 199–204, doi: 10.1109/ICACSIS51025.2020.9263162.
71. Sagar, K.; Saha, A. The effect of user variables on academic websites usability: An empirical study, *Int. j. stat. manag. syst.*, vol. 22, no. 2, Feb. 2019, pp. 161–186, doi: 10.1080/09720510.2019.1580898.
72. Sagar, K.; Saha, A. Exploring the Effect of Tasks Difficulty on Usability Scores of Academic Websites Computed Using SUS, *Adv. Intell. Syst.*, vol. 1087, 2020, pp. 11–19, doi: 10.1007/978-981-15-1286-5_2/COVER.
73. Dianat, I.; Adeli, P.; Jafarabadi, M. A.; Karimi, M. A. User-centred web design, usability and user satisfaction: The case of online banking websites in Iran, *Appl Ergon*, vol. 81, Nov. 2019, p. 102892, doi: 10.1016/J.APERGO.2019.102892.
74. Lazar, J., Allen, A., Kleinman, J., Malarkey, C.: What frustrates screen reader users on the web: a study of 100 blind users. *Int. J. Hum. Comput. Interact.* 2007,22(3), 247–269
75. Hasan, L.: Evaluating the usability of educational websites based on students' preferences of design characteristics. *Int. Arab J. e-Technol.*, 2014, 3(3), 179–193
76. Zhang, P., Dran, G.V., Blake, P., Pipithsuksunt, V.: A comparison of the most important website features in different domains: an empirical study of user perceptions. In: AMCIS 2000 Proceedings (2000)
77. Pearson, J.M., Pearson, A., Green, D.: Determining the importance of key criteria in web usability. *Manag. Res. News*, 2007, 30(11), 816–828
78. Tran, K. The Psychology Behind the Hiding Prices Online. https://www.omegatheme.com/blog/366-the-psychology-behind- the-hiding-prices-online, 2020.
79. Xue, X.; Poonia, M.; Abdulsahib, G.M.; Bajaj, R.K.; Khalaf, O.I.; Dhumras, H.; Shukla, V. On Cohesive Fuzzy Sets, Operations and Properties with Applications in Electromagnetic Signals and Solar Activities. Symmetry 2023, 15, 595. https://doi.org/10.3390/sym15030595
80. Xe, X.; Marappan, R.; Raju, S.K.; Raghavan, R.; Rajan, R.; Khalaf, O.I.; Abdulsahib, G.M. Modelling and Analysis of Hybrid Transformation for Lossless Big Medical Image Compression. Bioengineering 2023, 10, 333. https://doi.org/10.3390/bioengineering10030333
81. Xue, X.; Chinnaperumal, S.; Abdulsahib, G.M.; Manyam, R.R.; Marappan, R.; Raju, S.K.; Khalaf, O.I. Design and Analysis of a Deep Learning Ensemble Framework Model for the Detection of COVID-19 and Pneumonia Using Large-Scale CT Scan and X-ray Image Datasets. Bioengineering 2023, 10, 363. https://doi.org/10.3390/bioengineering10030363