IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 11, 2019, accepted December 1, 2019, date of publication December 23, 2019, date of current version January 14, 2020.

*Digital Object Identifier 10.1109/ACCESS.2019.2961768*

# Towards the Detection of UX Smells: The Support of Visualizations

**PAOLO BUONO, DANILO CAIVANO, MARIA FRANCESCA COSTABILE, (Senior Member, IEEE), GIUSEPPE DESOLDA, AND ROSA LANZILOTTI**
Computer Science Department, University of Bari Aldo Moro, 70125 Bari, Italy
Corresponding author: Giuseppe Desolda (giuseppe.desolda@uniba.it)

This work was supported in part by the Italian Ministry of Education, University and Research through the RPASinAir Project under Grant ARS01_00820, and in part by the Apulia Region through the Auriga2020 Project under Grant T5LXK18.

**ABSTRACT** Daily experiences in working with various types of computer systems show that, despite the offered functionalities, users have many difficulties, which affect their overall User eXperience (UX). The UX focus is on aesthetics, emotions and social involvement, but usability has a great influence on UX. Usability evaluation is acknowledged as a fundamental activity of the entire development process in software practices. Research in Human-Computer Interaction has proposed methods and tools to support usability evaluation. However, when performing an evaluation study, novice evaluators still have difficulties to identify usability problems and to understand their causes: they would need easier to use and possibly automated tools. This article describes four visualization techniques whose aim is to support the work of evaluators when performing usability tests to evaluate websites. Specifically, they help detect ''usability smells'', i.e. hints on web pages that might present usability problems, by visualizing the paths followed by the test participants when navigating in a website to perform a test task. A user study with 15 participants compared the four techniques and revealed that the proposed visualizations have the potential to be valuable tools for novice usability evaluators. These first results should push researchers towards the development of further tools that are capable to support the detection of other types of UX smells in the evaluation of computer systems and that can be translated into common industry practices.

**INDEX TERMS** User interfaces, website evaluation, design for quality, user eXperience, usability.

## I. INTRODUCTION AND MOTIVATION

Nowadays, computer systems are widely available, but only some of them are capable to provide a positive User eXperience (UX) to people interacting with them. UX extends the more traditional concept of usability, focused primarily on ease of learning and ease-of-use, and emphasizes aspects like aesthetics, emotions and social involvement, which have a major impact on the pleasure and satisfaction of people using a computer system [1], [2]. UX is still a broadly defined term; a universal definition of UX does not yet exist, but it is acknowledged that the experience of a user with a software product is influenced by functional quality attributes of the product (e.g. utility, robustness), by non-functional quality attributes (e.g. usability, privacy) and by specific UX attributes (e.g. desirability, pleasure) [1].

The efforts of addressing UX in the development of computer systems keep growing, with the aim of creating systems able to better motivate, engage and satisfy their users. However, daily experiences in working with various software artefacts (including websites of different types) show that, despite the powerful functionality they offer, people have many difficulties in using them [3]. Indeed, a user interface hard to understand and use causes many problems to users. Usability is the extent to which users are able to use a system, product or service to perform their activities in their specific context of use with effectiveness, efficiency and satisfaction [4]. It is well known that a low level of usability means that users cannot work out how to use a system, and this compromises a lot the overall UX (see, e.g. [2]). Thus, usability is a system quality attribute that greatly affects UX.

The literature proposes different methods and tools that support designers and evaluators in addressing usability. Since the 80ths the research on Human-Computer Interaction (HCI) has proposed the Human-Centred Design (HCD)

The associate editor coordinating the review of this manuscript and approving it for publication was Zhihan Lv.

as the design and development model, currently described in the ISO 9241 - Part 210, that has to be followed in order create usable interactive systems. A basic principle of HCD is to analyze the users' perspective of the system to be designed, the tasks to be performed with the system and the context of use. The core of this model is to develop the system by iterating a design-implementation-evaluation cycle, and the evaluation, possibly involving users, represents the key activity.

The HCI research has also provided principles and guidelines that can drive designers in taking their decisions for ensuring usability. Applying design guidelines is a good starting point, but there are no ''cookbooks''; therefore, there are not alternatives to system evaluation. Various methods can be used for evaluating systems at the different phases of their development. Still today such methods are seldom applied in the actual software development practice for several reasons [3], [5]. This happens primarily because i) developers think that usability evaluation methods are very much resource-demanding; ii) they do not have enough expertise in usability evaluation; iii) usability evaluation is scarcely automated. Thus, there is the need to adequately train developers in performing usability evaluation as well as to create tools that can support novice evaluators in both identifying usability problems and understanding their causes, so that these problems can be fixed.

A large part of our research work deals with e-government websites, specifically websites of small municipalities in Italy, which are designed and managed by web editorial staff by using Content Management Systems [6]. These websites offer general information (e.g., addresses of public offices, their work hours and contacts), and provide online services to citizens, with the potential of drastically reducing costs and waiting times. Only 37% of Italian internet users interact with Public Administration (PA) websites [7]. In addition to an incomplete digitisation of PA services, this situation stems from a lack of trust in websites: too often people find them incomplete, outdated, and especially affected by usability problems, such as difficult to understand content because of incoherent page layouts and intricate navigation paths. The Department of the Public Administration of the Italian Government is determined to improve the quality of PA websites. To this goal, it has set up a Working Group on Usability,[1] which devotes special attention to websites of small municipalities (the majority in Italy). Indeed, big PA organizations have the resources to properly address usability by involving experienced website designers and evaluators. Small municipalities, as those analyzed in [8], can only offer websites developed by web editorial staff, who have very little or no experience on usability and UX evaluation. Novice evaluators need methods that are simple to apply and require limited resources in terms of time and people, but also any other possible tools that may help them in usability and UX evaluation.

Software tools that support usability evaluation studies already exist and, in many cases, perform some preliminary analyses based on *metrics*, such as *task success* or *time spent to perform a task*. However, evaluators must still devote a big effort to identify further problems and to understand their causes. Thus, they could benefit from the identification of the so-called *usability smells*, which are defined as indicators of software interface aspects of low usability that may hamper users in correctly accomplishing tasks [9]. The term ''code smells'' is very popular in the Software Engineering Community. As introduced by Fowler [10], it indicates pieces of code of lower quality, which violate some fundamental design principles. Code smells are not bugs nor technically incorrect code structures. They indicate weaknesses in the code structure that may favor the emergence of some kind of problems in the future. Code smells are strictly related to technical debt [11], i.e. the amount of cost/effort of additional rework needed, in order to improve code structure, and thus removing code smells.

In [12], usability smells indicate possible problems that need refactoring, i.e., they need changes to the navigation, presentation or business processes of a web application with the purpose of improving its usability. Recently, some authors have being working on tools that automatically gather interaction data from users, detect usability smells and propose a concrete solution in terms of usability refactoring [9], [13]–[16]. In this article, we present a new semi-automatic approach to detect usability smells in website navigation. Specifically, four visualizations are presented that aim to support usability evaluators identifying usability smells by showing the paths followed by users that interacted with a website to execute tasks of a usability test; the optimal path to successfully complete a task (the one requiring the minimum number of actions by the user) is also visualized, in order to help pointing out web pages that confuse participants and lead users to follow wrong paths. We opted for this approach since we believe that fully automatic tools are useful for the initial processing of the data collected during a study, but the role of evaluators is instrumental in analysing problematic web pages based on the usability smells revealed by the visualizations. Only evaluators can deepen the real causes of the difficulties users encounter during the interaction with a website.

The proposed visualizations apply and customize existing visualization techniques, which are here used with the novel purpose of providing usability smells to evaluators. The results of the study reported in this article show that these visualizations may provide a valid support to evaluators in detecting certain types of usability smells. As already remarked, usability dimensions are also important UX dimensions; thus, usability smells can be considered types of UX smells. We believe that these results will encourage researchers to go further towards the definition of other types of UX smell, proposing new methods and tools that can be translated into common industry practices.

[1] http://www.funzionepubblica.gov.it/glu

The article is organised as follows. Related work is discussed in Section 2. Section 3 presents the visualization techniques. Section 4 reports an experimental study that compares the four techniques, in order to investigate how they support usability evaluators and to highlight their differences in terms of evaluator performance and satisfaction. Section 5 concludes the article.

## II. RELATED WORK

The related work is organized along with two subsections. The first one is about tools that support usability testing of software systems, in particular websites. The second section addresses the importance of usability evaluation of e-Government websites.

### A. TOOLS FOR USABILITY TESTING

Various methods have been developed for evaluating usability of interactive systems. A common distinction is between methods that involve users (user-based methods) and analytical methods [17]. User-based methods include user testing: usability is assessed by requiring that a sample of real users use the system to perform specific tasks. Some authors actually consider user testing as the most reliable method, because only when users interact with the systems, we can make sure of the problems they face [3]. In a usability test, the evaluators identify specific tasks and ask users to perform these tasks to the best of their ability, while evaluators watch them, possibly listen to them, and take notes. Data such as audio/video recordings, evaluators' notes, PC logs, results of questionnaire administered to test participants are gathered and analyzed, in order to detect usability issues. Despite the significant information it provides, usability testing is often neglected since it is considered expensive, especially due to the evaluators' expertise it requires and to the time necessary for data analysis [5], [18].

Several tools are available to assist the different phases of remote or local usability testing. Examples are Morae [19], Ovo Solo [20], Validately [21], UTassistant [22], Userlytics [23]. Even if these tools speed-up the test design and execution, automatize data storage, and help perform some analysis, it is the evaluator, based on his/her expertise, that must understand usability problems and their causes.

Other tools are available in the market, whose main goal is to provide information about the user interaction with web sites, such as new and returning visitors, referrals, bounce rates and real-time visitor statistics. Examples are Adobe Analytics, Chartbeat, Google Analytics, Matomo, Zap, to name but a few. Typically, they automatically gather data on user behavior and visualize the results on pie charts, histograms, timelines, etc. Even if these visual representations help understanding user behavior, they do not provide explicit indications about usability issues.

In some cases, the heatmap visualization is exploited to highlight the web page areas that users mainly interacted with [24], [25]. The heatmap is a visualization overlayed on a web page to show where people clicked/touched, moved the mouse, scrolled [25]; by analyzing the heatmap, the evaluator can see the areas of the web page that mostly attracted the users' attention. However, heatmap does not directly provide any specific indication about usability problems. Actually, it is very time-consuming for the evaluator to look at the heatmap of each web page and at its areas and/or widgets; in addition, the fact that certain areas mostly attract the attention of users does not imply that these areas present usability problems.

Some research works provided different solutions to drive evaluators in finding usability issues. For example, in [26] the user's actions during a task and the optimal actions to perform a task are shown on a timeline, in order to help the evaluator to see the user's actions and the deviations from the correct ones, speeding up the time to investigate something that occurred at a certain time. Similarly, *WUP* is a tool for remote usability evaluation of websites supporting evaluators in carrying out a visual comparison between actual user behaviour and an optimal sequence of actions visualized on an interactive timeline [27]. *WebQuilt* visualizes data from interaction log files using a Node-Link representation, whose nodes are the web pages visited by users, links indicate the transition from one page to another and their thickness indicates the frequency of users' transitions [28]. Although WebQuilt is the most similar technique to the ones presented in this article, it has a low level of interactivity, limited to scrolling and zooming the different timelines. In addition, there is no evidence about the real benefits since it was not evaluated in a real context. Lettner *et al.* [29] proposed an automatic approach to find usability issues avoiding user tests: data logs reporting users interactions carried out on a given device or website are analysed to find similarity with predefined tasks (e.g., writing an email, buy an item). All the interactions grouped for each task are visualized using Sankey Diagrams, enriched with histograms, to facilitate the analysis of user navigation paths. Despite the data display is very clear and compact, the information provided is only related to transitions between predefined and high-level tasks: it does not permit to analyze the user behaviour in more details or to highlight usability critical issues.

The literature also reports the so-called *clickstream* analysis tools [30]–[32], which visualize data about users' behavior. Actually, they address very big data, referring to millions of users, and the visualizations are quite complex. They are not of interest in this research, where the analysis is limited to the navigation paths of the participants to a usability test, which in most studies ranges from 10 to 50 people.

### B. USABILITY OF E-GOVERNMENT SITES

e-Government refers to the use of Information and Communication Technologies (ICT), particularly web-based applications, to provide faster, cheaper, easier, and more efficient access to and delivery of information/services to citizens, businesses, other agencies (non-profit), and governmental entities. Even if the literature reports on the advantages of e-Government (e.g. cost savings), the use of e-Government services is still very limited.

Focusing on the European Union, member states are at different levels of development and use of e-Government services. Denmark, Sweden, the Netherlands are among the leaders not only in Europe but also in the world, while several countries, especially in Eastern Europe, are progressing more slowly. A composite index, called Digital Economy and Society Index (DESI), summarizes relevant indicators on Europe's digital performance and tracks the evolution of EU member states in digital competitiveness [33]. DESI includes 5 main dimensions:

1) *Connectivity*, which measures the deployment of broadband infrastructure and its quality;
2) *Human Capital*, which measures the skills needed to take advantage of the possibilities offered by a digital society;
3) *Use of Internet*, which accounts for the variety of activities performed by citizens online;
4) *Integration of Digital Technology*, which measures the digitization of businesses and their exploitation of the online sales channel;
5) *Digital Public Services*, which measures the digitization of public services, and focuses in particular on e-Government and e-Health.

In DESI 2019, Finland, Sweden, the Netherlands, and Denmark scored the highest ratings; they are among the global leaders in digitalization. These countries are followed by the United Kingdom, Luxembourg, Ireland, Estonia, and Belgium. Other countries still have a long way to go, and the EU needs improvements by all countries to be able to compete on the global stage. Italy ranks 24th out of the 28 EU Member States. More specifically, Italy performs relatively well, although still below the EU average, as regards *Connectivity* and *Digital Public Services*. However, three out of ten people are not regular internet users yet, and more than half of the population lacks basic digital skills. This situation also derives from the low usability of the PA websites, with hard-to-understand content, inconsistent page layout, intricate navigation paths, hard-to-find help [34].

In order to improve this situation, PA websites should be designed by taking into great account usability and UX aspects. As mentioned in the introduction, the Department of the Public Administration of the Italian Government has created a Working Group on Usability, which has been primarily focused on developing tools that support PA website staff in design and evaluation activities (see http://www.funzionepubblica.gov.it/glu). One of the valuable results of the Working Group is the eGLU LG 2018.1 Protocol, which guides PA web staff (that generally has very little or no expertise in usability evaluation) in performing simplified but still effective usability testing of the websites they work on [35]. This Protocol is a practical guide that explains how to organize and perform usability tests; it also provides all the needed artifacts (e.g. modules to report the usability problems, spreadsheets to analyze the acquired data according to the identified measurement criteria) and describes the specific steps to organize and perform a usability test by using the thinking aloud technique, which is acknowledged as instrumental to perform cost-effective evaluations (see e.g., [17]).

More recently, as a result of the work within a project sponsored by the Italian Ministry of Economic Development (MISE), a web-based platform has been developed, aimed at supporting remote usability testing according to eGLU LG 2018.1 evaluation protocol. This platform, called eGLU Box PA, provides Italian PAs with a lightweight and simple service that does not require any installation on user devices and has no other special requirements (e.g., specific operating system and/or processors). Experiments conducted with eGLU Box PA have shown that it provides great support to PA organizations in designing and running usability testing [36]. eGLU Box PA is in two versions, Italian and English, so that it can be used worldwide. Similarly to other tools available in the market (e.g. Morae [19], Ovo Solo [20]), eGLU Box PA helps to perform some analyses on the collected data. For example, it computes metrics, such as the task success rate (i.e., the percentage of tasks that users correctly complete during the test), that are visualized in tables. Based on the mouse/keyboard logs, it also provides heatmaps of the visited web pages.

The success of such automatic tools indicates that usability evaluators get valuable support from them. In most cases, these tools can reveal tasks whose execution created difficulties for users. However, it is not easy to identify in which pages, along the user navigation path, the problems occurred. To this, evaluators should perform a video-analysis of the test sessions to identify web pages that confuse users. This is very resource-demanding and requires evaluator's experience. In order to ease this analysis, we designed and developed the visualization techniques reported in this article.

## III. IDENTIFICATION OF USABILITY SMELLS

This section describes four interactive visualization techniques whose common aim is to reveal usability smells in website navigation, which are important hints on web pages possibly affected by usability issues. Specifically, each technique visualizes the paths followed by all participants of a usability test when performing a specific task. Thus, if N tasks were performed in the test, N visualizations are produced, one for each task. To give an example, let us consider a usability test in which M participants performed N tasks. N visualizations are produced, each one showing the navigation paths of the M participants while performing that task. The optimal path to perform the task is also shown, since deviations from this path provide a good indication to the evaluator about the pages to be further analysed, which possibly present usability issues.

The visualizations are based on four commonly used techniques to visualize website navigation paths or, in general, navigations in graph-based structures (representing e.g. streets, social networks, etc.): *Arc Diagram* [37], [38], *Word tree* [39], *Sankey Diagram* [40], [41], and *Node-Link* [42]–[44]. The web pages are represented as nodes and user transitions from one page to another are represented as links. Node label reports the title of the web page and each link has a weight that indicates the number of users that performed the transition. The visualizations are interactive, allowing evaluators to perform basic interactions such as show/hide labels of the nodes (web pages), move nodes and links, perform.

The visualizations use the same visual encoding, namely:

- The blue node is the *starting* page of the task;
- The green node is the *target* page, i.e. the page where the user successfully ends the task;
- Black nodes are *transit* nodes, i.e. pages visited by users to reach the target page;
- Red nodes are web pages where users wrongly terminated the task;
- Green links show transitions that are part of the optimal path;
- Grey links show transitions that are not part of the optimal path;
- Violet links show backward transitions, namely, transitions to a previously visited page;
- Node size is proportional to the number of users who visited the corresponding web page;
- Link thickness is proportional to the link weight, i.e. the number of transitions the link represents.

As shown in the rest of this section, the visualizations reveal usability smells through their elements, e.g.:

*a*) Outgoing transitions from a node on the optimal *path*. If from a node A on the optimal path a large number of transitions start, going to nodes that are not on the optimal path, it is very likely that some widgets of the web page associated with node A confused the users, leading them to follow a wrong path, for example by clicking on a link with a not appropriate label.

*b*) Red nodes. Since these nodes indicate pages where the task terminated incorrectly, it could mean that 1) users that finished their task on a red node wrongly considered the related page as the right one for completing the task, or 2) users got lost during task execution, in particular during the visit of the last pages before reaching the page represented by the red node.

*c*) Backward paths. If users come back following a backward transition starting from a node A, it means that A has not been considered the right page, thus a navigation error can occur on the page(s) that wrongly pushed users in going to A.

In the following, each technique is described by illustrating a visualization that shows the paths followed by 15 users performing a task during a usability test; the test was executed in March 2019, in order to evaluate the Italian Navy website (www.marina.difesa.it). The task required users to get information about the San Marco Brigade.

To perform this task, the user should visit the pages as in the following:

1. The user starts from the homepage, in which many images are shown, some of them are even clickable. However, in order to perform the task in the fastest way, the user must click on the burger menu at the top right of the page. Once open, this menu shows six items and one of this is "The Fleet", that the user must click to open "The Fleet" page.
2. "The Fleet" page shows several clickable images, one of these has the label "Special forces". The burger menu is still available at the top right of the page. By clicking on the "Special Forces" label, the linked page appears.
3. The "Special Forces" page shows other clickable images, while the top bar with the burger menu is still present. The user must click on the image with label "San Marco Brigade" to reach the page providing information about this brigade.
4. Having reached the "San Marco Brigade" page, the task is successfully concluded.

The ideal path to correctly accomplish this task thus requires the visit of 4 pages and is composed of 3 links, connecting the homepage node to the San Marco Brigade node. Notice that the Italian Navy website has been recently updated and might not reflect the described path.

An important remark is that scalability is not an issue, since the visualization techniques are used to support usability evaluators with low experience in analysing the paths followed by the participants of a usability test, who navigated in a website to perform the test tasks. Such tests usually involve a number of participants in the range 10-50 and require the execution of not very complex tasks.

### A. ARC DIAGRAM VISUALIZATION

The first visualization, shown in Figure 1, is based on Arc Diagram [37], [38]. It places nodes (web pages) along a horizontal line and connects them with arcs (links). In order to reduce arc length and visual cluttering, nodes are shown from left to right according to the following: the starting node is the first one on the left and the target node is the last one on the right; for each node $i$ the successive one is the node $j$, whose transition $(i, j)$ is the one performed by the highest number of users among all transitions starting from $i$. The visualization in Figure 1 shows 10 outgoing transitions from the starting page (the blue node) to the one labelled "The Fleet" along the optimal path (the connecting arc is green). From this page, 13 transitions reached a web page of the optimal path ("Special Forces") and finally 10 transitions correctly reached the "San Marco Brigade" target page (the green node). Some users followed wrong paths or wrongly terminated their tasks. For example, from the starting node there are also 5 transitions to a page that is not on the optimal path ("We Are Navy"), thus 1/3 of theoutgoing transitions from the home page are wrong, indicating a potential usability issue on this page. Another noteworthy page is the one corresponding to the node labeled "Special Forces", which has 13 incoming transitions but only 10 outgoing transitions.

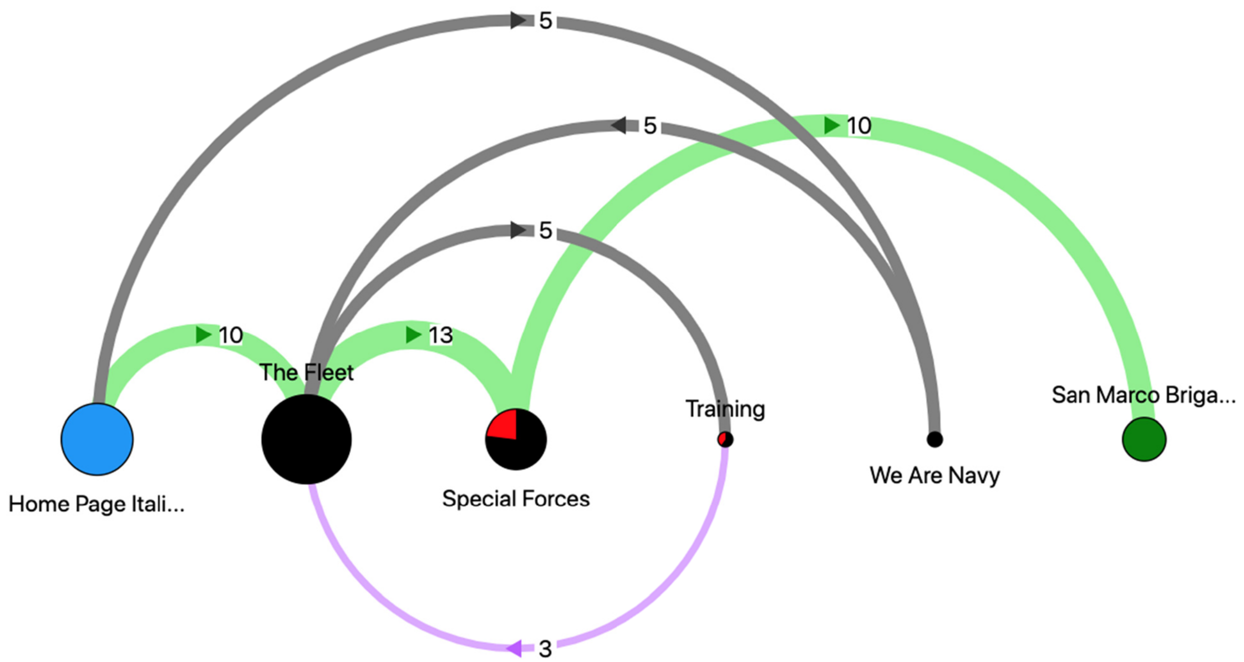


**FIGURE 1.** The Arc Diagram visualization of the paths of the 15 users among six webpages. Only 10 users completed the task, arriving to the green target node. Arc diagram does not permit people to see how many users followed the optimal path. Other 3 users stopped the task at the page "Special Forces" and 2 users stopped at the page "Training". The arc in purple shows 3 backward transitions.

It is represented as a red-black pie chart where the red slice reveals that 3 users wrongly terminated their task at this page (slice details are revealed on mouseover). Similarly, the page "Training" was wrongly considered as target by 2 users, while 3 users came back to "The Fleet" page, as indicated by the purple arc.

Notice that the arcs that start from a node have weights whose sum is not always equal to the number of test participants due to the backward arcs. For example, the 13+5 transitions that start from the "The Fleet" node are more than the involved 15 participants, since at that node 3 backward transitions arrive from the "Training" page. Moreover, being a website structured as a graph, when the number of participants and transitions between pages is high, the order of nodes may not always satisfy the rule that the incoming node is on the right of the outgoing node (see for example the transition from "We Are Navy" to "The Fleet" in Figure 1). In order to avoid any confusion, each arc contains a small arrow that indicates the navigation direction.

Node labels are shown below or above the node, alternatively, to improve the visualization readability. If a label is longer than a certain number of characters, it is cut and ellipsis are added at the end (see "San Marco Briga…" in Figure 1).

### B. PAGE TREE VISUALIZATION

The second visualization is called Page Tree since it is inspired by Word tree that, in its original version, shows 'words in context' to support text analysis in long documents [39]. Looking at Figure 2, we see that 10 transitions correctly occurred from the starting page to the "The Fleet" page (the transition are green), from which there are 4 outgoing transitions along the optimal path, 2 still along the optimal path but terminating the task on a wrong intermediate page("Special Forces"), and 3+1 along two other paths.

In this visualization, links always go from left to right. If a link $(i, j)$ represents a transition that goes back to a previously visited node $j$, this node $j$ is shown again to the right of node $i$. The fact that a node may appear multiple times makes more difficult to understand the total number of visited web pages.

A positive aspect of this visualization is that it is the only one that clearly shows the number of users that completed the task by following the optimal path. Indeed, the weights on the links of the green path going from the starting node to the "San Marco Briga…" node at the top right of Figure 2 clearly indicates that 4 users followed the optimal path. The target node was also reached by other 6 people, 2 following another path and 4 following a third path. This technique well represents the branched structure of different choices made by users. On purpose, the nodes where users stopped the task are aligned on the right. This design choice may be modified in the future.

### C. SANKEY DIAGRAM VISUALIZATION

The third visualization is based on the Sankey Diagram technique [40, 41], also called flow diagram because it emphasizes data flow through consecutive states, represented as nodes. Nodes are shown as bars whose height is proportional to the number of visits to the related page. Links are represented as bands whose width is proportional to link weight, i.e. to the number of users performing that transition.

Figure 3 clearly shows that a total of 10 users arrived to the target page. However, as in Arc Diagram, it is not

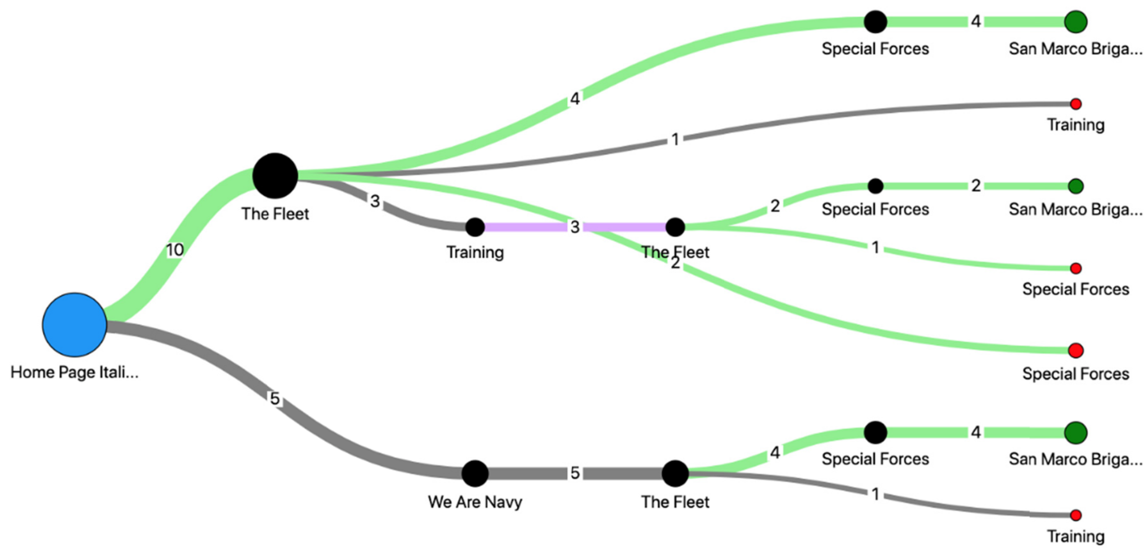


**FIGURE 2.** Page Tree visualization of the paths of the 15 users performing the same task as in Figure 1. Some nodes may appear multiple times: this makes the analysis a bit more difficult, but the number of users that moved along a specific path is clearly shown. It is easy to see that 4 users reached the target page following the optimal path, while other 6 (2+4) followed other paths.

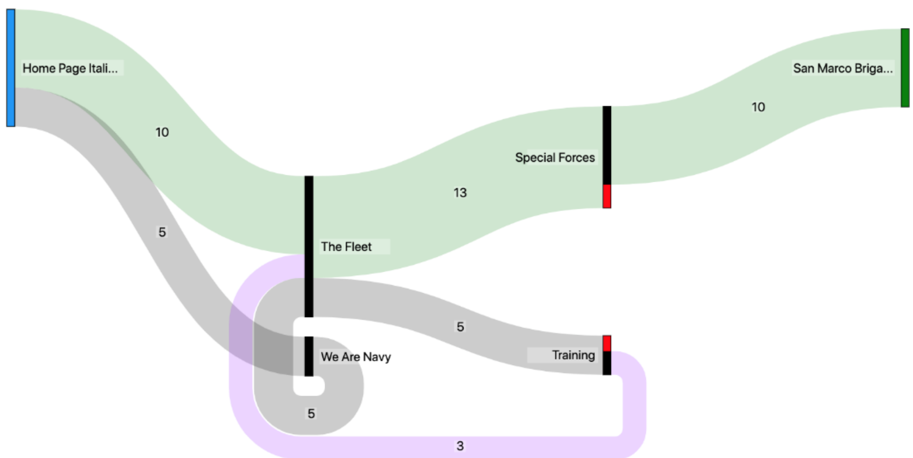


**FIGURE 3.** Sankey Diagram visualization of the paths of the 15 users performing the same task as in Figure 1 and 2. Nodes are represented by bars that may also have more colors (see "Special Forces" and "Training"). Links are represented as bands whose width is proportional to the number of users that performed that transition.

possible to see how many people followed the optimal path. The "Special Forces" and "Training" nodes are shown as a bar with two colors. In the case of "Special Forces", the red color reveals that 3 users wrongly terminated their task, while the black color indicates that 10 users proceeded towards the target node. In the case of the "Training" node, 2 users wrongly terminated their task, while the backward band (in purple) from "Training" to "The Fleet" indicates that 3 users went back to the previous page. Notice that visualizations with many backward paths could be confusing due to the overlapping bands.

### D. NODE-LINK VISUALIZATION

The last visualization, shown in Figure 4, is based on the traditional Node-Link visualization, the most common for a website (e.g., see [42]–[44]). The deviations from the optimal path are immediately visible, e.g., the link with weight 5 from "Homepage Marina Militare" to "We Are Navy" or the 3 backward link from the "Training" node to "The Fleet".

The Node-Link visualization has well-documented problems, e.g. link crossings and links with different slopes increase the visual complexity and the readability of the transitions. Ghoniem et al. show that even for a simple task such as locating a node or finding the links between two nodes, node-link diagram performs badly, even for graphs with as few as 20 nodes [45].

Interaction mechanisms, which allow the user to move nodes and to highlight connections, partially alleviate the problems at the cost of shifting the visual interpretation from the user perceptive (pre-attentive) system to the cognitive system.

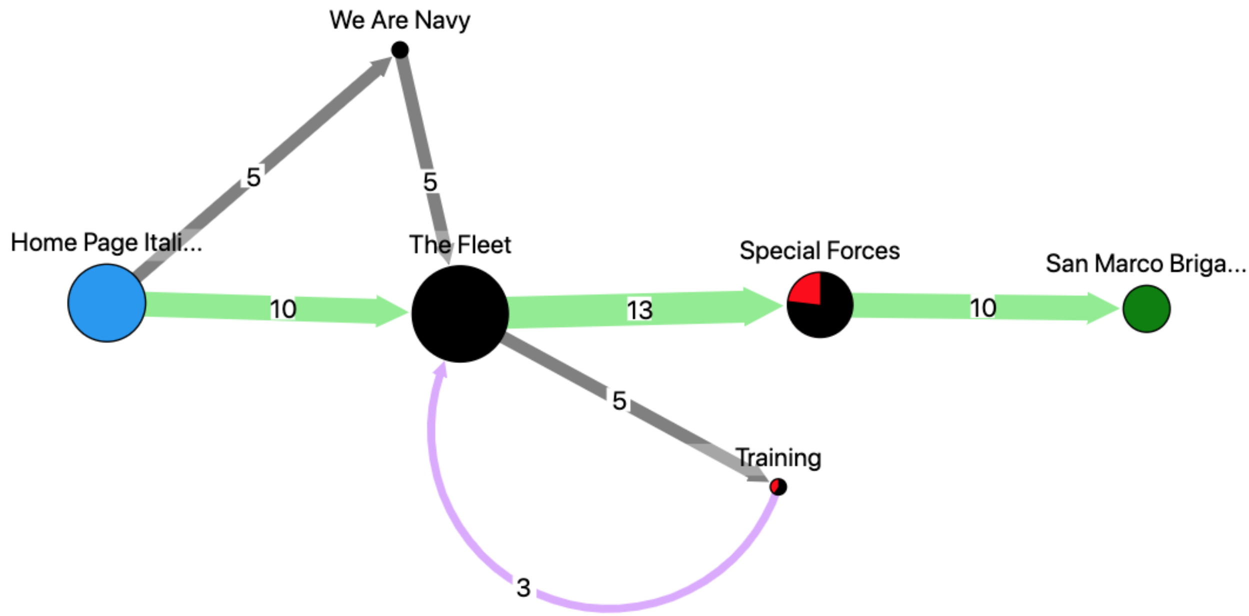


**FIGURE 4.** Node-Link visualization of the paths of the 15 users performing the same task as in Figure 1-3.

This might make the analysis potentially slower and more tiring.

## IV. EXPERIMENTAL STUDY

The visualization techniques have been developed by following an HCD approach in which prototypes of increasing complexity were created and analyzed in formative evaluation sessions; the results of such sessions were instrumental to improve the quality of the successive prototypes. The first evaluations were rather informal involving three HCI experts who inspected the prototypes independently. Then, in peer-review sessions their comments were discussed together and taken into account to develop the new prototypes. User tests with thinking aloud protocol were also performed involving 12 different users, 3 for each visualization.

By using the visualization techniques in the current version described in Section 3, an experimental study was performed in order to better understand the value of these techniques for novice evaluators in detecting usability smells. More specifically, the user study was designed to answer the following research questions:

**RQ1.** Are the visualizations techniques able to support evaluators in identifying usability smells in the web pages?

**RQ2.** What is the difference among the visualization techniques in terms of evaluator satisfaction?

### A. PARTICIPANTS AND STUDY DESIGN

We recruited 15 participants (11 males, 4 females) among the students of the third year of Computer Science, who are novice evaluators since they completed a course on Human-Computer Interaction, acquiring basic experience on usability evaluation, in particular on user testing. Their average age was 23.86 years (min= 23, max=32, SD= 2.28). A within-subject design was performed with visualization technique as an independent variable and four within-subject levels, namely the four visualizations Arc Diagram, Page Tree, Sankey Diagram, and Node-Link.

### B. THE EXPERIMENTAL TASKS

The participants were told to behave as usability evaluators that analysed the data collected in a recent usability test in which employees of an Italian Public Administration interacted with the official website of the Italian Navy. Specifically, participants had to look at each one of the four visualizations and perform the five tasks listed below, which required to identify specific pages and users' navigation paths.

The 5 tasks were defined by the authors of this paper. Each participant repeated the tasks 4 times (in a different order), each time looking at one of the 4 visualizations of the navigation paths relative to the same task from the Navy usability test; specifically, the visualizations used in the test are those shown in Figures 1-4.

The 5 experimental tasks were:

T1. Locate the path(s) that led to the task failure;
T2. Identify the alternative paths that led to task success;
T3. Indicate the number/percentage of users who followed alternative paths without being able to complete the task;
T4. Identify the web page most visited by users (but the homepage);
T5. Identify the web page that mostly confused the users.

Because each one of the 15 participants executed the above 5 tasks on 4 visualizations, the total number of trials was 300 ($15 \times 5 \times 4$).

### C. PROCEDURE

The study took place in a quiet university room where the study apparatus was installed. A laptop with a 17-inch display provided with an external mouse was available. Two HCI experts were involved: one acted as observer, the other as facilitator. The study lasted 3 days: 5 participants were individually observed each day. All participants followed the same procedure. First, they were introduced to the study purpose and what they had to do. Nobody refused to take part in the study; they were rewarded with a 8Gb USB memory stick. Participants were asked to sign a consent form. After that, they filled in a questionnaire for collecting their demographic data, competences on IT, and experience in performing usability testing.

The participants were provided with a booklet composed of 4 pages: each page referred to one of the four visualization techniques and reported the tasks to be performed with that specific technique. To avoid carry-over effect, the experimental tasks on each page, as well as the booklet pages, were ordered so that the visualization technique order was counterbalanced across the participants, and the task order was counterbalanced across the experimental conditions, both according to a Latin Square design.

The facilitator introduced and explained the first visualization, i.e., the one reported on the first page of the participant booklet. Then, the participant read aloud the first task text and started to interact with the visualization. At the end of all the experimental tasks, the participant filled in an online questionnaire including SUS and NASA-TLX surveys to express his/her satisfaction with the technique they had just used. For each visualization technique, the time taken for tasks execution plus questionnaire filling is less than 10 minutes. Before repeating the same procedure with the next technique, the participant was invited to relax for five minutes.

This procedure was preliminarily assessed by a pilot study with three participants different from those involved in the experimental sample, in order to check the overall research methodology (e.g., time constraints, coding techniques, video-recording activities).

### D. DATA COLLECTION AND ANALYSIS

Quantitative and qualitative data were collected in order to answer the two research questions. All the interactions were audio-video recorded by using a camera. Two researchers transcribed the observer's notes and the questionnaire open questions. They also analyzed and wrote out some useful parts of the audio-video recordings. Inductive Thematic Analysis was carried out on these data [46]. Then, the two researchers independently double-checked the results. The initial reliability value was 79%, thus the researchers discussed the differences and reached a full agreement.

To analyze the *support* provided by the visualizations to evaluators in detecting usability smells, metrics such as the *task success rate* and *task execution time* were considered. The two researchers created an excel file reporting for each task performed by each user the following data: user ID (from 1 to 15), visualization technique name, task ID (T1-T5), task success rate (i.e. the percentage of participants that successfully completed the task with respect to the total number of participants), task execution time (in seconds). Then, they independently double-checked such data. The initial reliability value was 91%, thus the researchers discussed the differences and reached a full agreement.

The online questionnaire to investigate *satisfaction* with the visualization technique was composed of two well-known questionnaires: 1) the System Usability Scale (SUS) to estimate tool usability as perceived by the user [47], and 2) the NASA Task Load Index (NASA-TLX) used as "Raw TLX", to assess the workload caused by each visualization [48], because the workload the user feels when using a software tool has an influence on user satisfaction. Appendix A reports a short description of the two questionnaires.

One-way repeated measures ANOVAs (all Greenhouse-Geisser corrected) with post-hoc pairwise comparisons (Bonferroni corrected) were adopted to assess significant differences among the four visualization techniques for task execution time, success rate, SUS score and NASA-TLX index. Shapiro-Wilk test has been used to verify that all involved data were normally distributed.

### E. RESULTS AND DISCUSSION

The first research question is related to the *support* the proposed visualizations provided to evaluators in identifying usability smells in website navigation. The support was assessed in terms of task execution *time* and task *success rate*. Tables 1–5 show descriptive statistics as well as inferential ones used for analysing the 5 tasks executed by the evaluators during the study. We can see that the success rate is pretty high in all 5 tasks; this is a first indication that the visualizations were able to support the evaluators' analysis. This section presents in details and discusses all the obtained results. It is worth remarking that quantitative and qualitative data were triangulated for a deeper investigation of the results. Triangulated data are discussed in the following only when they provide further insights.

Task T1 requires the identification of paths that led to task failure; these paths are usability smells related to web pages with potential usability issues. Table 1 shows that, while the success rate was the maximum for all visualizations, significant differences emerged in terms of time. Specifically, the post-hoc analysis revealed that Sankey Diagram and Node-Link allowed evaluators to locate the path/s ending with task failure faster than Page Tree, while no difference emerged with respect to Arc Diagram.

Concerning the qualitative data, 5 evaluators said that "*in Sankey Diagram the ramifications of a node are clearly visible and easy to interpret*", thus supporting a quick and easy identification of incorrect paths. Similar positive comments were externalized by 4 evaluators for Node-Link. Page Tree takes more time possibly due to a larger number of ramifications. Four evaluators stated that Page Tree "*creates some confusion due to the high number of nodes and links*".

**TABLE 1.** Task execution time and task success rate for task T1.

| | Visualization | Time | | Success rate | |
|---|---|---|---|---|---|
| | | x̄ | SD | x̄ | SD |
| Task T1 | Sankey D. | 5.58 | 2.61 | 100 | 0 |
| | Arc D. | 10.29 | 6.29 | 100 | 0 |
| | Page Tree | 12.07 | 6.74 | 100 | 0 |
| | Node-Link | 5.75 | 2.85 | 100 | 0 |
| | *Test* | $F(2.031, 862.84)$ =6.75 $p$ = .004 | | N.A. | |
| | *Post-hoc* | Sankey D. – Page ($p$=.050) Node – Page ($p$=.041) | | N.A. | |

**TABLE 2.** Task execution time and task success rate for task T2.

| | Visualization | Time | | Success rate | |
|---|---|---|---|---|---|
| | | x̄ | SD | x̄ | SD |
| Task T2 | Sankey D. | 41.20 | 23.84 | 96.67 | 12.90 |
| | Arc D. | 29.33 | 3.88 | 93.33 | 17.59 |
| | Page Tree | 53.70 | 32.47 | 83.33 | 24.39 |
| | Node-Link | 62.44 | 33.08 | 63.33 | 39.94 |
| | *Test* | $F(1.948, 17183.65)$ =3.293 $p$ = .062 | | $F(1.640, 26000)$ =4.846 $p$ = .023 | |
| | *Post-hoc* | N.A. | | Sankey D. – Node ($p$=.042) | |

Another problem highlighted by 3 evaluators concerns the backward paths that, although present in Page Tree, create some confusions since "*the direction of all the links goes from left to right, thus it is confusing to see the backward paths that follow the same direction of the forward paths while indicating an opposite browsing direction*". In Arc Diagram, "*the overlapping of the arcs and their alternate directions makes slower the identification of incorrect paths*", as noted by 3 evaluators. Despite these differences and difficulties related to efficiency, by using each visualization techniques the evaluators were able to correctly detect the paths that led to task failure, without significant differences.

Task T2 requires the identification of alternative paths that led to reach the target page. This is a usability smell that helps evaluators identify web pages that deviated users from the ideal path, even if they successfully reached the target page. As shown in Table 2, no statistically significant difference emerged among the four techniques concerning the time spent by the evaluators. The ANOVA only revealed a significant difference among the success rates. In particular, the post-hoc test highlights that Sankey Diagram was better than Node-Link. A possible explanation of this result can be the spatial arrangement of the nodes and the link overlapping in Node-Link, which could cause an incorrect identification of the alternative paths. Indeed, 6 evaluators remarked that *"Node-link has an unorganized layout due to overlapped arcs"*, thus tasks like path identification are more error prone. The complex layout of visualizations based on node-link is a well-known problem: existing algorithms can only optimize criteria like the intersections of nodes [49]–[51] or link [52], [53] but it is impossible to always avoid intersections of nodes or arcs. In graph theory, the optimization of arc (link) intersection is known as "planarity testing" problem, i.e., the algorithm problem of verifying if a graph can be represented in the plane as a planar graph without arc intersections. Despite the algorithm implemented in our Node-Link optimizes link intersection, this visualization sometimes might generate errors in path identification due to arc overlapping, as it happened for task T2.

Task T3 requires evaluators to identify the number/percentage of users who followed alternative paths without being able to complete the task. In this case, the usability smell is related to web pages that confused users and pushed them in following wrong paths.

As shown in Table 3, the ANOVA test found statistical differences among the visualizations, and the post-hoc analysis highlighted that Arc Diagram was better than Page Tree in terms of time. The evaluators commented positively on both Sankey Diagram and Arc Diagram. For example, 5 evaluators stated that "*the ramifications of a node are clearly visible*" in Sankey Diagram, while 4 evaluators said that with Arc Diagram "*it is very easy to identify the number of participants who followed alternative paths counting the arc values*". Even if Arc Diagram required more time in identifying specific paths (Task T1), it outperforms all the visualizations in identifying the number/percentage of users who followed alternative paths. A possible explanation of this result is its linear and close arrangement of nodes and arc, which facilitates the counting of the number of participants following a given path to reach specific nodes. With Page Tree, we observed that evaluators have the additional load of summing the number of arcs related to the same path but distributed in different parts of the visualization. Despite there is not a significant statistical difference, it is evident that participants were quite slow to complete this task by using Node-Link; a possible explanation is the arc overlapping discussed for Task T2, since we noticed that participants sometimes needed manual reorganization of nodes to complete tasks like this one, thus requiring more time than other techniques, but without lowering its effectiveness.

Task T4 requires evaluators to identify the most visited web page. This reveals a usability smell if, for example, a large number of incoming transitions arrive to a wrong node. No difference emerged for this task in terms success rate, while a difference emerged for the time (see Table 4). Post-hoc test did not reveal significant differences in pairwise comparison. The only notable result is that the Arc Diagram allowed the evaluators to identify the most visited page faster than Node-Link with a p-value slightly greater than the 0.5 threshold ($p$ =.081).

Similar to Task T3, the analysis of participants interactions revealed that they were faster in accomplish this task with

**TABLE 3.** Task execution time and task success rate for task T3.

| | Visualization | Time | | Success rate | |
|---|---|---|---|---|---|
| | | x̄ | SD | x̄ | SD |
| Task T3 | Sankey D. | 13.07 | 13.39 | 100 | 0 |
| | Arc D. | 12.33 | 10.32 | 100 | 0 |
| | Page Tree | 22.50 | 3.59 | 100 | 0 |
| | Node-Link | 21.80 | 14.37 | 100 | 0 |
| | *Test* | $F(2.075, 5437.69)$ =3.469 $p$ = .043 | | N.A. | |
| | *Post-hoc* | Arc – Page ($p$=.005) | | N.A. | |

**TABLE 4.** Task execution time and task success rate for task T4.

| | Visualization | Time | | Success rate | |
|---|---|---|---|---|---|
| | | x̄ | SD | x̄ | SD |
| Task T4 | Sankey D. | 12.41 | 6.79 | 100 | 0 |
| | Arc D. | 9.82 | 4.15 | 93.33 | 25.82 |
| | Page Tree | 14.50 | 7.24 | 100 | 0 |
| | Node-Link | 20.93 | 13.09 | 93.33 | 25.82 |
| | *Test* | $F(2.013, 2853.11)$ =4.311 $p$ = .024 | | $F(1.730, 14333)$ =.651 $p$ = .509 | |
| | *Post-hoc* | N.A. | | N.A. | |

**TABLE 5.** Task execution time and task success rate for task T5.

| | Visualization | Time | | Success rate | |
|---|---|---|---|---|---|
| | | x̄ | SD | x̄ | SD |
| Task T5 | Sankey D. | 12.16 | 8.21 | 73.33 | 45.77 |
| | Arc D. | 26.90 | 19.94 | 80.00 | 41.40 |
| | Page Tree | 26.63 | 25.15 | 73.33 | 45.77 |
| | Node-Link | 31.77 | 21.10 | 73.33 | 45.77 |
| | *Test* | $F(2.074, 11474.12)$ =2.064 $p$ = .151 | | $F(1.249, 37000)$ =.189 $p$ = .724 | |
| | *Post-hoc* | N.A. | | N.A. | |

Arc Diagram since the linear and close arrangement of the nodes along a one-dimensional axis facilitates node visual comparison. The other visualizations distribute the nodes in a 2D plane, thus making more complex and slower their size comparison.

Task T5 asks evaluators to identify the web page that created most problems. Table 5 shows that there are no statistically differences in terms of time and success rate. However, mean and standard deviation of the time spent by participants as well as the evaluators' comments externalized during and after the test, indicate that Sankey Diagram was considered the technique that better supports evaluators in identifying the web page that created most problems.

Concerning the second research question (RQ2), namely if there are differences among the four visualization techniques in terms of evaluator *satisfaction*, SUS score and NASA-TLX were analyzed (see Table 6). No differences emerged for both of them. The SUS scores of the four visualizations are positive

**TABLE 6.** SUS score and NASA-TLX workload for each visualization.

| Visualizations | SUS score | | NASA workload | |
|---|---|---|---|---|
| | x̄ | SD | x̄ | SD |
| Sankey D. | 77.03 | 19.17 | 20.33 | 7.10 |
| Arc D. | 67.53 | 21.98 | 23.66 | 9.86 |
| Page Tree | 69.44 | 19.43 | 21.00 | 8.97 |
| Node-Link | 66.36 | 18.40 | 22.33 | 9.89 |
| Test | $F(2.590,1483.083)$ = .615 $p$ = .583 | | $F(2.510, 93.33)$ = .741 $p$ = .512 | |
| Post-hoc | N.A. | | N.A. | |

results, according to the findings reported in [54]. This is an encouraging result that confirms our design choices. The absence of difference among these scores is understood as an upward level of satisfaction, without negative or exceptionally positive cases. This good result was confirmed by the low NASA-TLX workload for all the four visualizations. The low score of the NASA-TLX shows that participants worked easily. However, a study with a larger number of participants could reveal differences that might now be hidden by a high variance of the results.

The conclusion of the above analysis and discussion is that there is not a leading technique since each of them resulted adequate to identify usability smells. In general, Sankey Diagram was the visualization most appreciated by the evaluators. Thanks to the clear structure of navigation paths, Sankey Diagram supports a fast location of paths that led to task failure (T1), and an accurate identification of alternative paths that led to task success (T2). The immediate and clear structure of the navigation paths resulted as a positive aspect also of Node-Link, which allows evaluators to quickly identify paths that led to task failure (T1). However, the arc overlapping problem and the unorganized layout of Node-Link cause mistakes when evaluators identify alternative paths that led to task success (T2) or the page most visited by users (T4). Arc Diagram results adequate for the identification of different usability smells. Indeed, evaluators were faster when calculating the number of users who followed alternative paths without being able to complete the task (T3) and when they identified the page most visited by users (T4). These positive performances are mainly given by the one-dimensional arrangement of nodes that facilitates the identification of such kinds of usability smells. No positive performances or evaluators comments emerged for Page Tree that, despite it has been appreciated for a clear path ramification, obtained low performances in detecting usability smells, mainly due to the redundancy of nodes and arcs.

## V. CONCLUSION

This article reported how existing visualization techniques may be used with a novel valuable purpose, i.e., to support novice evaluators to detect a specific type of UX smells, namely usability smells in website navigation, during the analysis of the data collected in a test for website evaluation.

**TABLE 7. The SUS questionnaire.**

| | Strongly disagree | | | | Strongly agree |
|---|---|---|---|---|---|
| 1. I think that I would like to use this system frequently. | 1 | 2 | 3 | 4 | 5 |
| 2. I found the system unnecessarily complex. | 1 | 2 | 3 | 4 | 5 |
| 3. I thought the system was easy to use. | 1 | 2 | 3 | 4 | 5 |
| 4. I think that I would need the support of a technical person to be able to use this system. | 1 | 2 | 3 | 4 | 5 |
| 5. I found the various functions in this system were well integrated. | 1 | 2 | 3 | 4 | 5 |
| 6. I thought there was too much inconsistency in this system. | 1 | 2 | 3 | 4 | 5 |
| 7. I would imagine that most people would learn to use this system very quickly. | 1 | 2 | 3 | 4 | 5 |
| 8. I found the system very cumbersome to use. | 1 | 2 | 3 | 4 | 5 |
| 9. I felt very confident using the system. | 1 | 2 | 3 | 4 | 5 |
| 10. I needed to learn a lot of things before I could get going with this system. | 1 | 2 | 3 | 4 | 5 |

More specifically, the presented techniques, called Arc Diagram, Page Tree, Sankey Diagram, and Node-Link after the visualization techniques they apply and customize, show the navigation paths followed by users when performing a test task, as well as the optimal path to successfully complete the task. An experimental study was carried out to understand the support of these visualization techniques to novice evaluators in identifying usability smells. The study showed that the four techniques are capable to support the identifications of usability smells; Sankey Diagram and Arc Diagram slightly outperformed the other techniques.

The study was also instrumental to highlight some modifications to perform on the current visualizations; for example, we are considering to use web page thumbnails instead of nodes, as proposed in [43], for easier identification of the web pages. In addition, since in all the visualization techniques, but Page Tree, the nodes may indicate web pages that are visited multiple times, we are working on a new colour coding that may also indicate this.

We are aware of the limitations of this first experimental study, due to the number and type of evaluators involved, the low complexity of the navigation paths shown by the visualizations, and the small number of administered tasks. Indeed, its objective was to highlight the potentialities of the visualization techniques, rather than comparing the four techniques in order to select the best one. To overcome these limitations, we are going to perform a number of further experiments that involve a larger number of evaluators with different expertise (PA web staff, IT companies employees) and websites of different levels of complexity. The collected empirical evidence will allow us to better highlight the value of the different visualizations and to indicate the types of UX smells they can reveal.

We are confident that the work presented in this article will push researchers towards the development of tools that may support the detection of other types of UX smells in the evaluation of computer systems. The described work is a first step of a more straightforward research stream whose ultimate goal is to identify the relevant smell types that can impact on the different dimensions of the UX and, at the same time, propose easy to use techniques and supporting tools for allowing software companies to perform a "sustainable" UX evaluation, from both resources and required expertise point of views. In line with the actual trend observed in the software community, we want to move towards an agile and lean approach to UX evaluation, which also exploits automated tools. In this way we hope that the research results will be more easily translated into common industry practices. Another milestone of this research path, also borrowing the results already obtained by other research communities, will be the definition of the concept of "UX debt", which might be used in the future by companies for quantifying and addressing the investments needed for the improvement of the UX in a computer system.

## APPENDIX SYSTEM USABILITY SCORE (SUS) AND NASA TASK LOAD INDEX (NASA-TLX)

This appendix reports the SUS and NASA-TLX questionnaires administered to the participants in the experimental study.

The System Usability Score (SUS) gives a measure of the users' subjective usability evaluation of a given system. It is a closed-ended questionnaire encompassing 10 statements on an ordinal 5-point Likert scale from "strongly disagree" to "strongly agree", as shown in Table 7. This questionnaire was chosen for its reliability, brevity and wide adoption [55].

The ten items have opposite polarity. For odd items (1,3,5,7,9), the most positive response is 5; for even items (2,4,6,8,10), the most positive response is 1. For each returned questionnaire, SUS provides a single score, which ranges from 0 to 100, calculated as it follows:

1. For odd items (1,3,5,7,9): subtract 1 from the user response.
2. For even items (2,4,6,8,10): subtract the user response from 5. In this way, all values scale from 0 to 4 (with 4 being the most positive response).
3. Add up the converted responses for each questionnaire (obtaining a value in the range 0-40) and multiply that total by 2.5, thus obtaining a value in the range 0-100.

The final score of the SUS is obtained by averaging the scores of all questionnaires; it indicates the overall participants' satisfaction about the system. In [56], a conventional threshold of 70 is proposed to consider acceptable the overall usability of the evaluated system.

The NASA-TLX estimates the workload users perceive while they are performing a task or immediately afterwards [48]. The NASA-TLX version used in this research is the "Raw TLX". As shown in Figure 5, it is a 6-item survey that rates the perceived workload in using a system through 6 subjective dimensions: Mental Demand, Physical Demand, Tem-

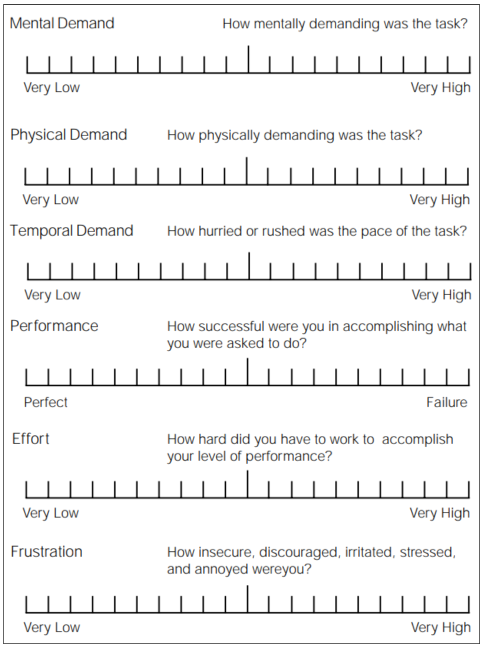


**FIGURE 5.** **The NASA-TLX questionnaire.**

poral Demand, Performance, Effort and Frustration. Each dimension gets a score in the range 5-100, since it is is rated through a scale of 20 values, being each value of 5 points. The user's workload is the average of the scores of the 6 dimensions. The overall workload is obtained by averaging all users' workloads [48].

## REFERENCES

[1] M. Hassenzahl and N. Tractinsky, "User experience—A research agenda," *Behav. Inf. Technol.*, vol. 25, no. 2, pp. 91–97, Mar. 2006.
[2] E. L.-C. Law, V. Roto, M. Hassenzahl, A. P. Vermeeren, and J. Kort, "Understanding, scoping and defining user experience: A survey approach," in *Proc. 27th Int. Conf. Hum. Factors Comput. Syst. (CHI)*. New York, NY, USA: ACM, 2009, pp. 719–728.
[3] C. Ardito, P. Buono, D. Caivano, M. F. Costabile, and R. Lanzilotti, "Investigating and promoting UX practice in industry: An experimental study," *Int. J. Hum.-Comput. Stud.*, vol. 72, no. 6, pp. 542–551, Jun. 2014.
[4] *Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts*. Standard ISO/IEC 9241-11, 2018.
[5] A. Fernandez, E. Insfran, and S. Abrahão, "Usability evaluation methods for the Web: A systematic mapping study," *Inf. Softw. Technol.*, vol. 53, no. 8, pp. 789–817, Aug. 2011.
[6] S. Federici, M. L. Mele, M. Bracalenti, A. Buttafuoco, R. Lanzilotti, and G. Desolda, "Bio-behavioral and self-report user experience evaluation of a usability assessment platform (UTAssistant)," in *Proc. Int. Conf. Hum. Comput. Interact. Theory Appl. (HUCAPP)*, 2019.
[7] *European Commission—Digital Single Market*. Accessed: Jun. 10, 2019. [Online]. Available: https://ec.europa.eu/digital-single-market/en/scoreboard/italy
[8] R. Lanzilotti and M. F. Costabile, "Usability evaluation of local government Websites: Lessons learnt," Dept. Comput. Sci., Public Admin., IVU, Univ. Bari Aldo Moro, Rome, Italy, Tech. Rep. 4/2017, Nov. 2017.
[9] J. Grigera, A. Garrido, J. M. Rivero, and G. Rossi, "Automatic detection of usability smells in Web applications," *Int. J. Hum.-Comput. Stud.*, vol. 97, pp. 129–148, Jan. 2017.
[10] M. Fowler, *Refactoring: Improving the Design of Existing Code*. Reading, MA, USA: Addison-Wesley, 2018.
[11] H. L. O'Brien, P. Cairns, and M. Hall, "A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form," *Int. J. Hum.-Comput. Stud.*, vol. 112, pp. 28–39, Apr. 2018.
[12] A. Garrido, G. Rossi, and D. Distante, "Refactoring for usability in Web applications," *IEEE Softw.*, vol. 28, no. 3, pp. 60–67, May 2011.
[13] V. Lelli, A. Blouin, B. Baudry, F. Coulon, and O. Beaudoux, "Automatic detection of GUI design smells: The case of Blob listener," in *Proc. 8th ACM SIGCHI Symp. Eng. Interact. Comput. Syst. (EICS)*. New York, NY, USA: ACM, 2016, pp. 263–274.
[14] J. Grigera, A. Garrido, and J. M. Rivero, "A tool for detecting bad usability smells in an automatic way," in *Proc. Int. Conf. Web Eng. (ICWE)*. Cham, Switzerland: Springer, 2014, pp. 490–493.
[15] F. Paternò, A. G. Schiavone, and A. Conte, "Customizable automatic detection of bad usability smells in mobile accessed Web applications," in *Proc. 19th Int. Conf. Hum.-Comput. Interact. Mobile Devices Services (MobileHCI)*. New York, NY, USA: ACM, 2017, Art. no. 42.
[16] J. Grigera, A. Garrido, and G. Rossi, "Kobold: Web usability as a service," in *Proc. IEEE/ACM Int. Conf. Automated Softw. Eng. (ASE)*. Piscataway, NJ, USA: IEEE Press, Oct./Nov. 2017, pp. 990–995.
[17] J. Preece, H. Sharp, and Y. Rogers, *Interaction Design: Beyond Human—Computer Interaction*. Hoboken, NJ, USA: Wiley, 2015.
[18] J. Nielsen and H. Loranger, *Prioritizing Web Usability*. London, U.K.: Pearson, 2006.
[19] *Morae*. Accessed: Jun. 10, 2019. [Online]. Available: https://www.techsmith.com/morae.html
[20] *OVO Solo*. Accessed: Jun. 10, 2019. [Online]. Available: http://www.ovostudios.com/ovosolo.asp
[21] *Validately*. Accessed: Jun. 10, 2019. [Online]. Available: https://validately.com/
[22] S. Federici, M. L. Mele, R. Lanzilotti, G. Desolda, M. Bracalenti, F. Meloni, G. Gaudino, A. Cocco, and M. Amendola, "UX evaluation design of UTAssistant: A new usability testing support tool for Italian public administrations," in *Human-Computer Interaction. Theories, Methods, and Human Issues* (Lecture Notes in Computer Science), M. Kurosu, Ed. Cham, Switzerland: Springer, 2018, pp. 55–67.
[23] *Userlytics*. Accessed: Jun. 10, 2019. [Online]. Available: https://www.userlytics.com/
[24] E. Arroyo, T. Selker, and W. Wei, "Usability tool for analysis of Web designs using mouse tracks," in *Proc. Extended Abstr. Hum. Factors Comput. Syst. (CHI EA)*. New York, NY, USA: ACM, 2006, pp. 484–489.
[25] V. Navalpakkam and E. Churchill, "Mouse tracking: Measuring and predicting users' experience of Web-based content," in *Proc. ACM Annu. Conf. Hum. Factors Comput. Syst. (CHI)*. New York, NY, USA: ACM, 2012, pp. 2963–2972.
[26] F. Paternò, A. G. Schiavone, and P. Pitardi, "Timelines for mobile Web usability evaluation," in *Proc. Int. Working Conf. Adv. Vis. Interfaces (AVI)*. New York, NY, USA: ACM, 2016, pp. 88–91.
[27] T. Carta, F. Paternò, and V. F. de Santana, "Web usability probe: A tool for supporting remote usability evaluation of Web sites," in *Proc. IFIP Conf. Hum.-Comput. Interact. (INTERACT)*. Berlin, Germany: Springer, 2011, pp. 349–357.
[28] S. J. Waterson, J. I. Hong, T. Sohn, J. A. Landay, J. Heer, and T. Matthews, "What did they do? Understanding clickstreams with the WebQuilt visualization system," in *Proc. Int. Conf. Adv. Vis. Interfaces (AVI)*, 2002, pp. 94–102.
[29] F. Lettner, C. Grossauer, and C. Holzmann, "Mobile interaction analysis: Towards a novel concept for interaction sequence mining," in *Proc. 16th Int. Conf. Hum.-Comput. Interact. Mobile Devices Services (MobileHCI)*. New York, NY, USA: ACM, 2014, pp. 359–368.
[30] B. Albert, T. Tullis, and D. Tedesco, *Data Preparation*, B. Albert, T. Tullis, and D. Tedesco, Eds. San Mateo, CA, USA: Morgan Kaufmann, 2010, ch. 5, pp. 107–119.
[31] J. Lee, M. Podlaseck, E. Schonberg, and R. Hoch, "Visualization and analysis of clickstream data of online stores for understanding Web merchandising," *Data Mining Knowl. Discovery*, vol. 5, pp. 59–84, Jan. 2001.
[32] R. Kateja, A. Rohith, P. Kumar, and R. Sinha, "VizClick visualizing clickstream data," in *Proc. Int. Conf. Inf. Vis. Theory Appl. (IVAPP)*, Jan. 2014, pp. 247–255.
[33] *European Commission—DESI Index*. Accessed: Sep. 19, 2019. [Online]. Available: https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi
[34] E. Buie and D. Murray, Eds., *Usability in Government Systems: User Experience Design for Citizens and Public Servants*. Amsterdam, The Netherlands: Elsevier, 2012.
[35] *eGLU LG 2018.1 Protocol*. Accessed: Sep. 19, 2019. [Online]. Available: http://www.funzionepubblica.gov.it/glu

[36] S. Federici, M. L. Mele, R. Lanzilotti, G. Desolda, M. Bracalenti, A. Buttafuoco, G. Gaudino, A. Cocco, M. Amendola, and E. Simonetti, ''Heuristic evaluation of eGLU-box: A semi-automatic usability evaluation tool for public administrations,'' in *Proc. Int. Conf. Hum.-Comput. Interact. (HCII)*. Cham, Switzerland: Springer, 2019, pp. 75–86.
[37] P. Neumann, S. Schlechtweg, and S. Carpendale, ''ArcTrees: Visualizing relations in hierarchical data,'' in *Proc. Eurovis (EGEV)*, 2005, pp. 53–60.
[38] M. Wattenberg, ''Arc diagrams: Visualizing structure in strings,'' in *Proc. IEEE Symp. Inf. Vis. (INFOVIS)*, Jun. 2002, pp. 110–116.
[39] M. Wattenberg and F. Viégas, ''The word tree, an interactive visual concordance,'' *IEEE Trans. Visual. Comput. Graph.*, vol. 14, no. 6, pp. 1221–1228, Nov. 2008.
[40] M. Sankey, ''Introductory note on the thermal efficiency of steam-engines,'' in *Proc. Minutes Inst. Civil Eng.*, vol. 134, 1898, pp. 278–283.
[41] P. Riehmann, M. Hanfler, and B. Froehlich, ''Interactive Sankey diagrams,'' in *Proc. IEEE Symp. Inf. Vis. (INFOVIS)*, Oct. 2005, pp. 233–240.
[42] Y. Wang, W. Dai, and Y. Yuan, ''Website browsing aid: A navigation graph-based recommendation system,'' *Decis. Support Syst.*, vol. 45, no. 3, pp. 387–400, Jun. 2008.
[43] C. Plaisant, J. Grosjean, and B. B. Bederson, ''SpaceTree: Supporting exploration in large node link tree, design evolution and empirical evaluation,'' in *Proc. IEEE Symp. Inf. Vis. (INFOVIS)*, Oct. 2002, pp. 57–64.
[44] J. Heer and D. Boyd, ''Vizster: Visualizing online social networks,'' in *Proc. IEEE Symp. Inf. Vis. (INFOVIS)*, Oct. 2005, pp. 32–39.
[45] M. Ghoniem, J.-D. Fekete, and P. Castagliola, ''On the readability of graphs using node–link and matrix–based representations: A controlled experiment and statistical analysis,'' *Inf. Vis.*, vol. 4, no. 2, pp. 114–135, Jun. 2005.
[46] V. Braun and V. Clarke, ''Using thematic analysis in psychology,'' *Qualitative Res. Psychol.*, vol. 3, no. 2, pp. 77–101, Jan. 2006.
[47] J. Brooke, ''SUS—A quick and dirty usability scale,'' *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.
[48] S. G. Hart, ''NASA–task load index (NASA–TLX); 20 years later,'' *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 50, no. 9, pp. 904–908, Oct. 2006.
[49] P. Eades and X. Lin, ''Spring algorithms and symmetry,'' *Theor. Comput. Sci.*, vol. 240, no. 2, pp. 379–405, Jun. 2000.
[50] X. Huang and W. Lai, ''Force-transfer: A new approach to removing overlapping nodes in graph layout,'' in *Proc. Australas. Comput. Sci. Conf. (ACSC)*, vol. 16. Darlinghurst, NSW, Australia: Australian Computer Society, 2003, pp. 349–358.
[51] W. Li, P. Eades, and N. Nikolov, ''Using spring algorithms to remove node overlapping,'' in *Proc. Asia–Pacific Symp. Inf. Vis. (APVis)*, vol. 45. Darlinghurst, NSW, Australia: Australian Computer Society, 2005, pp. 131–140.
[52] J. Hopcroft and R. Tarjan, ''Efficient planarity testing,'' *J. ACM*, vol. 21, no. 4, pp. 549–568, Oct. 1974.
[53] J. M. Boyer and W. J. Myrvold, ''On the cutting edge: Simplified O (n) planarity by edge addition,'' *J. Graph Algorithms Appl.*, vol. 8, no. 3, pp. 241–273, 2004.
[54] A. Bangor, P. Kortum, and J. Miller, ''Determining what individual SUS scores mean: Adding an adjective rating scale,'' *J. Usability Stud.*, vol. 4, no. 3, pp. 114–123, 2009.
[55] S. Borsci, S. Federici, and M. Lauriola, ''On the dimensionality of the system usability scale: A test of alternative measurement models,'' *Cogn. Process*, vol. 10, no. 3, pp. 193–197, Aug. 2009.
[56] A. Bangor, P. Kortum, and J. Miller, ''The system usability scale (SUS): An empirical evaluation,'' *Int. J. Hum.-Comput. Interact.*, vol. 24, no. 6, pp. 574–594, 2008.

**PAOLO BUONO** received the Ph.D. degree in computer science on visual data analysis. He is currently an Associate Professor with the Department of Computer Science, University of Bari Aldo Moro, Italy. He is also the Co-Ordinator of information visualization and visual analytics with the Interaction Visualization Usability (IVU) and UX Laboratory. His current research interests include visual analytics, information visualization, human–computer interaction, mobile applications, and time series.

**DANILO CAIVANO** is currently an Associate Professor of software engineering and project management with the Department of Computer Science, University of Bari Aldo Moro, and a Consultant for companies and organizations especially in the field of research and development projects. He is also the Head of the SERLAB Research Laboratory and the Director of the short master in cyber security. He contributed to the creation of The Hack Space, Cyber Security Laboratory, University of Bari. He is also a member of the Board of Director of the Southern Italy Chapter Project Management Institute, the Co-Ordinator of the PMI-SIC Academy, and a member of the Technical Scientific Committee of the Apulian Information Technology District and the IT Strategic Steering Committee.

**MARIA FRANCESCA COSTABILE** is currently a Full Professor with the Computer Science Department, University of Bari, Italy, where she teaches human–computer interaction and other courses for the computer science curricula. She is also the Director of the Interaction Visualization Usability (IVU) and UX Laboratory. Her current research interests are primarily on interaction design, usability engineering, UX, information visualization, and end-user development. She is also a member of ACM. In 2018, she received the ACM SIGCHI Lifetime Service Award. She started the Italian Chapter of ACM SIGCHI and has served as the First Chair, from 1996 to 2000.

**GIUSEPPE DESOLDA** is currently a Research Fellow with the Computer Science Department, University of Bari Aldo Moro, Italy. He is also a member of the Interaction Visualization Usability (IVU) and UX Laboratory, where he coordinates research on novel interaction techniques, the Internet of Things, and usable security. His research interests are on human–computer interaction and, in particular, interaction with the IoT and ubiquitous systems, linked open data, usability, and UX. He is a member of ACM, ACM SIGCHI, and SIGCHI Italy (the Italian Chapter of ACM SIGCHI). He is also a member of the Italian Working Group on Usability called Gruppo di Lavoro per l'Usabilità (GLU).

**ROSA LANZILOTTI** is currently an Associate Professor with the Computer Science Department, University of Bari Aldo Moro, Italy. She is a member of the Interaction Visualization Usability (IVU) and UX Laboratory. She coordinates research on usability engineering and UX. Her research interests include human–computer interaction and, in particular, usability and UX, accessibility, technology-enhanced learning, statistical methods for data analysis, end-user development, and UX practices in industry and public institutions. She is also a member of ACM, ACM SIGCHI, and SIGCHI Italy (the Italian Chapter of ACM SIGCHI). She is also a member of the Italian Working Group on Usability called Gruppo di Lavoro per l'Usabilità (GLU), set up by the Italian Ministry of Public Administration (PA), whose aim is to spread the culture of usability and user experience (UX) in PA.