Techset Composition Ltd, Salisbury
Doc: {IEE}SEN/Articles/Pagination/SEN20100062.3d

www.ietdl.org

Published in IET Software
Received on 28th May 2010
Revised on 20th May 2011
doi: 10.1049/iet-sen.2010.0062

ISSN 1751-8806

Empirical study to assess whether the use of routes facilitates the navigability of web information systems

V. de Castro[1] *M. Genero*[2] *E. Marcos*[1] *M. Piattni*[2]

[1]*Kybele Research Group, Rey Juan Carlos University, Tulipán S/N, Móstoles, Madrid 28933, Spain*
[2]*Alarcos Research Group, University of Castilla La Mancha, Paseo de la Universidad 4, Ciudad Real 13071, Spain*
E-mail: valeria.decastro@urjc.es

Abstract: The design of web information systems (WISs) often relies on a domain model and, at most, some informal guidelines, signifying that many design decisions which may directly affect usability are left to the designers. The study presented here has been carried out in order to analyse whether it is possible to limit this arbitrary aspect of the navigation design process. The authors have therefore carried out a controlled experiment and its replica to test whether it is effectively possible to obtain more easily navigable WISs by using a navigation model, including certain design characteristics, as a starting point. These characteristics are the use of a main menu which only contains the services required by the user and the signposting of the sequence of steps necessary to carry out each service correctly, denominated as a route. The empirical study carried out has permitted us to discover that, given two WISs with the same functionality and identical interface styles, users prefer a WIS with routes and a main menu including the beginning of each route, and perceive it to be more easily navigable. In addition, the use of routes in the navigation model improves the navigability of the WIS built from it, particularly for subjects with less experience, that in our case are 1st year Master students of Computer Science with 5 years in average using WISs and 1.5 years designing them. The principal benefit of the research is the identification of certain design principles which allow more navigable WISs to be built independently of web designers' skills.

1 Introduction

As the complexity of web information systems (WISs) grows, the difficulty involved in using their systems also grows. The users of traditional information systems usually spend a great deal of time on becoming familiar with the features and design of these systems. Conversely, on the web, users demand the ability to use a site on the basis of the web conventions that they have already acquired through their experience of using other sites, and they do not normally wish to read any manuals or instructions for individual sites [1, 2]. Several authors claim that a design oriented towards user needs is one of the key aspects of usability [1, 3]. 'Site design, is often more important for usability because users are never going to even get close to the correct pages unless the site is structured according to user needs and contains a navigation scheme that allows people to find what they want' [1]. Constantine and Lockwood [3] state that 'much software is designed and built with little consideration for how it will be used and how it can best support the work its user will be doing' and they also maintain that the model-driven design process is a key element for the usage-centred design.

Usability is undoubtedly one of the quality characteristics that most concerns the web engineering community. We are unable to mention the references to works dealing with usability here owing to space limitations. However, a summary of the main proposals concerning the evaluation of usability is presented in [4].

It is widely recognised in the web community that usability is significantly associated with navigation [1, 5, 6]. Good navigation aids permit users to acquire the information they are seeking quickly and efficiently, and therefore contribute to site users' perceived success [5]. Such navigability is usually expressed in web engineering methods through navigational models [7]. However, the majority of these methods focus on the analysis and design of navigation, leaving the final implementation of the navigation, and consequently certain aspects that could affect navigability, to the web designer's discretion.

We have identified certain design characteristics which, when applied to navigation modelling, allow WISs to be built, and which could, from our point of view, facilitate navigation through them [8]. These characteristics are the identification of 'routes' and 'a main menu including the beginning of each route'. A route is defined as 'the sequence of steps established for the WIS that the user must follow to execute a user service'. After implementing the routes in a WIS and signposting the sequence of steps that the user must follow to carry out each process correctly, the routes will guide users' navigation through the WIS. Drawing an analogy with the road system, the main menu

is analogous to a sign which indicates all possible destinations at a road junction, and the route is analogous to the signposting on the chosen road which indicates how to reach the chosen destination. Thus, in the same way as these characteristics on a road could help drivers to reach their destination, these characteristics in a WIS could help users to perform their required tasks.

However, although it might seem intuitively obvious to us that a WIS including signposted routes is more easily navigable, as Zelkowitz and Wallace point out, a new proposal in software engineering lacks credibility if there is no empirical evidence of its usefulness [9]. Conventional wisdom, intuition, conjectures and proofs of concepts are not therefore reliable sources of credible knowledge [10]. Experimentation is therefore essential if software engineering is to become a mature discipline. The realisation of families of experiments is thus fundamental if this situation is to be improved, given that isolated experiments rarely provide conclusive findings [10]. We have therefore carried out a family of experiments, consisting of an original experiment and its replica, to corroborate 'whether the use of routes and a main menu including the beginning of each route, really does improve the navigability of the WIS'. The improvement of navigability is here associated with the idea of improving usability from the navigation perspective, as is stated in [11]. This signifies that the usability of a WIS will be satisfactory if the navigation options allow the user to accomplish his/her task effectively, efficiently and easily.

In this empirical study we wish to compare the navigability of two WISs for conference organisation: ConfMaster [12], a well-known WIS in this field which is widely used in prestigious conferences; and WebConference [13], a similar WIS that we have built using the two new design characteristics. Both WISs have the same functionality and identical interface style, the only difference being in their navigation model, that is, the way in which users navigate the WIS: always starting at the beginning of each route, and navigating the route once a specific task has been selected. Therefore only the second WIS (WebConference) has routes and a main menu indicating the beginning of each route. Our reason for selecting this type of WIS for use in the experimentation lies in the fact that the application domain is very easy to understand, and is very simple to use and easy to explain to the students who participate in the experiments. Furthermore, despite its relative simplicity, it is of sufficient importance to permit us to undertake the measures required for the experiment.

The remainder of the paper is structured as follows. First, in Section 2 we analyse previous research works on the evaluation of usability in WISs. Section 3 describes the navigation model of the two WISs which are compared in the experiment, and the experiment's replica is described in Section 4. The analysis and interpretation of the empirical data are presented in Section 5. Section 6 discusses the implications of our work, analysing those kinds of web applications in which design characteristics such as those analysed in this work are of particular importance. Finally, Section 7 highlights the main contribution of the paper and future work.

2 Related work

User satisfaction is one of the most important measures of success in any information system, and particularly in WISs. Usability and navigation are two of the key factors in determining a user's satisfaction with a WIS [14]. Nielsen affirms that 'in the Web, usability is a necessary condition for survival' [1]. If a website is difficult to use, if the homepage fails to clearly state what a company offers and what users can do on the site or if users get lost on a website, they leave, because web users do not want to read a manual in order to use a particular website. The design of certain aspects that may affect a WIS' usability is, therefore, too costly a factor to be left to a website designer's discretion.

In recent years, and since the popularisation of WISs, various works analysing certain aspects that affect usability have appeared [6, 15–18]. Along with clarity of content and user relevance or motivation, ease of navigation (the aspect upon which this work is focused), is one of the most frequently referred to aspects with regard to usability.

Several works studying the effects of certain design characteristics of the web user interface on the ease of navigation exist, such as [1, 18–22]. A basic principle of human–computer interaction (HCI) is that user interface should be easy to use and predictable [18]. Works such as [1, 18, 19, 22] have studied the effect of the main menu on WIS, highlighting its importance as a determining success factor when designing usable WISs. Nielsen stresses the importance of the main menu, since it is a company's first contact with a customer, and is the means by which the company shows what it is offering and what users can do on the site [1]. Yu and Roh [19] presented an empirical study on the effect of different types of web menus (a simple selection menu, a global and local navigation menu, and a pull-down menu) on users. A similar work is presented in [22], in which the effect of different styles of menus is analysed. The navigational paths and how they affect user satisfaction is a further characteristic, and has been analysed by various authors such as [1, 18, 21]. Navigational paths are relevant as a means to prevent the user from getting lost and, for example, assist the user to discover where s/he is, what s/he is doing and where s/he has come from [1] or assist him/her to successfully carry out his/her tasks [18]. With regard to this, in [21] Gwizdka and Spence study the significance of metrics based on their similarity to an optimal path as good predictors of lostness and task success.

The aforementioned works therefore demonstrate that WISs can, to some extent, benefit from the applications of design principles that could improve their usability [1]. The main objective of our study is to investigate the effect of two particular design characteristics on WIS users: the routes (such as navigational paths) and a main menu including the beginning of each route.

Since ease of navigation is too generic a measurement, we decided to study measurements related to usability. ISO 9241-11 defines usability as 'the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use'. Fred Davis proposed two measurements related to satisfaction in a specified context of use: perceived usefulness and perceived ease of use. The first is defined as 'the degree to which a person believes that using a particular system would enhance his or her job performance' and the second is defined as 'the degree to which a person believes that using a particular system would be free from effort' [23]. Since we wish to measure the degree to which the WIS is 'easy' (free from effort) for the user to navigate, we decided to evaluate perceived ease of use.

We therefore eventually decided to measure the following three variables for our experiment in order evaluate the

doi: 10.1049/iet-sen.2010.0062

degree to which some particular design characteristics in a WIS improve its navigability:

- *Perceived ease of use*, defined as the degree to which a person believes that using a particular tool facilitates navigation through the WIS.
- *Effectiveness*, defined as how well the use of a particular WIS allows the required tasks to be accomplished. In our experiment the effectiveness is measured as the ratio between the number of clicks made and the number of clicks the task requires. The number of clicks required for each task is the number of mandatory steps required to finish an activity, and is different for each WIS.
- *Efficiency*, defined as the time spent in relation to how well the use of a particular WIS allows the required tasks to be accomplished. This variable was measured as the ratio between the number of correct clicks made and the time spent.

A more detailed description of how the data are collected and calculated is provided in Section 4, in which the experiment is presented.

With regard to the usability evaluation method used in this work, it was necessary to involve the users in the evaluation process. Various methods have been proposed for usability evaluation, and the most widely used of these are: usability inspection and user-based methods [24, 25]. Usability inspection methods involve expert evaluators who inspect the interface of an application in order to detect common defects or errors [26]. Common methods are cognitive walkthrough [27] and heuristic evaluation [26]. User-based methods, on the other hand, imply user participation in the evaluation. Representative real users are thus selected and observed when using the application, and their experiences and impressions are then collected by means of interviews or questionnaires [25].

In our experiment we decide to measure aspects related to the users' perception with regard to certain characteristics of the WIS, and we thus consider that a user-based method is most appropriate for our evaluation. While methods such as the use of a WIS and filling in a questionnaire allow as to obtain a real user's experience, usability inspection methods are usually based on identifying usability errors, and neither the users' effectiveness nor efficiency are explicitly measured [28].

The two WISs used in our experimentation are described below, and we shall then explain the experimental process followed to run the experiment and its replica.

3 WebConference against ConfMaster: similarities and differences

As was previously stated, the experiment compared two WISs: ConfMaster and WebConference. Both WISs have an identical interface style and offer the users (authors) the same functionality:

- The author can submit a paper.
- The author can modify her/his data.
- The author can view her/his submitted paper.

To carry out these tasks, an author must log into the system. Moreover, to submit a paper the author must register. The author can also ask for her/his password if s/he forgets it.

ConfMaster and WebConference WISs have an identical interface style, that is to say, the same colour and fonts, identical localisation of menus and titles and the same attributes in web form. Moreover, the data required to carry out the functionalities are the same in both WISs. The only difference between the two WISs considered in this experiment is in the design of their navigational model. According to [29] the navigation model comprises the navigation space (which objects can be visited in the navigation) and the navigation structure (how these objects are reached). In our case, differences in the navigation model are in the navigation structure: only WebConference WIS has signposted routes, and a main menu indicating the beginning of each route. Figs. 1 and 2 show the navigation models of the ConfMaster and WebConference WISs, respectively, while Fig. 3 shows screenshots of both WIS.

In this paper we compare both navigation models (ConfMaster and WebConference WISs) using the extended navigation model proposed by the MIDAS hypertext modelling method [8]. The extended navigation model is built as a UML model, which represents all the concepts

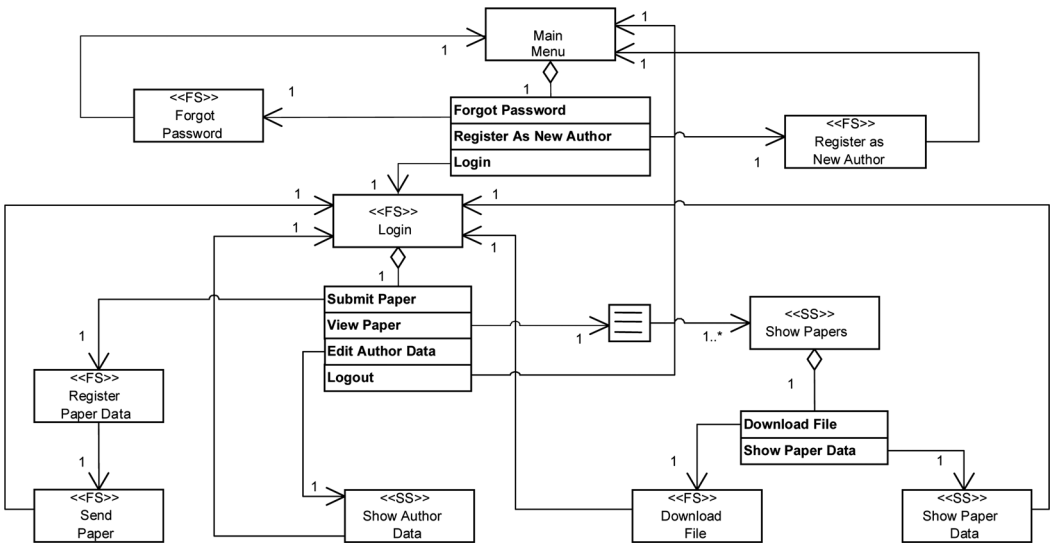


Fig. 1 *Extended navigation model of ConfMaster*

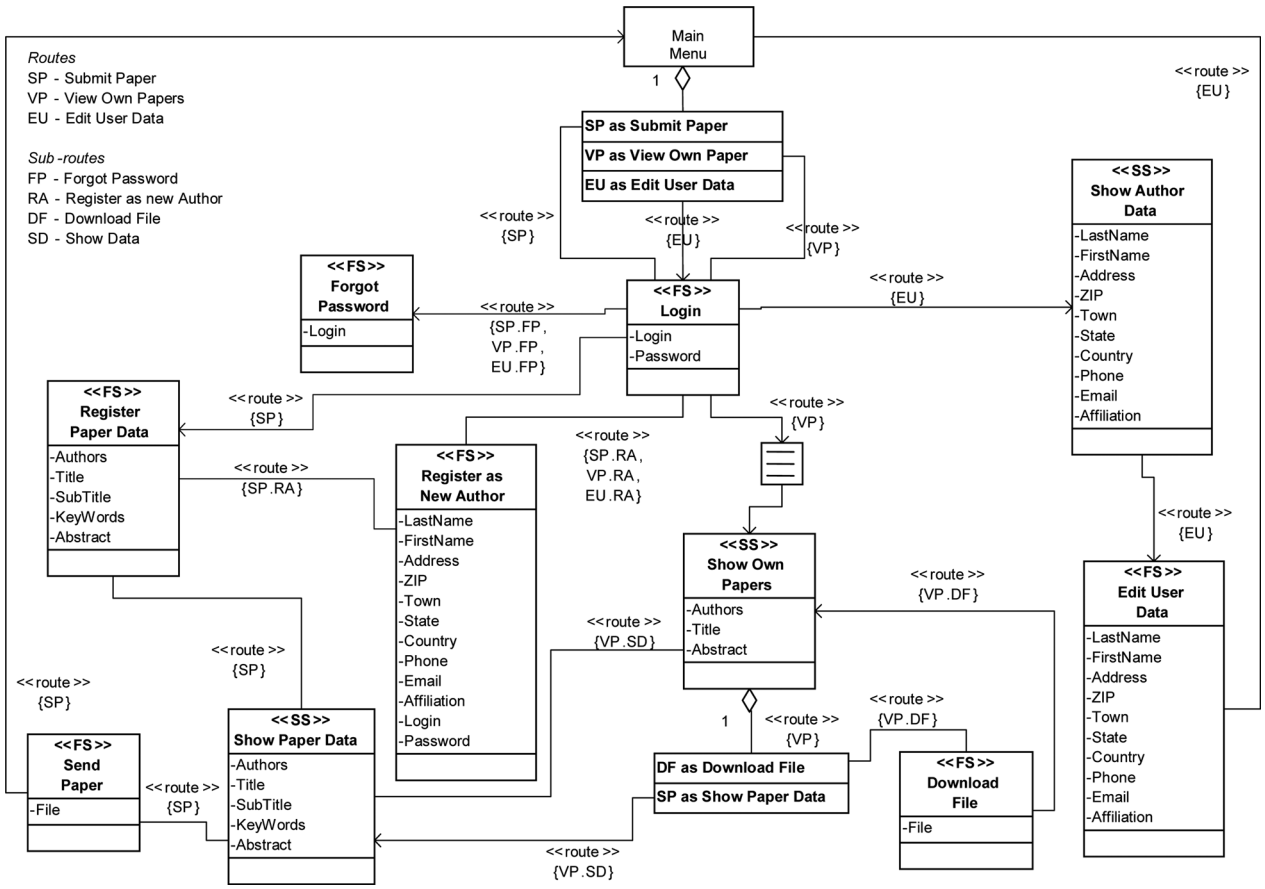
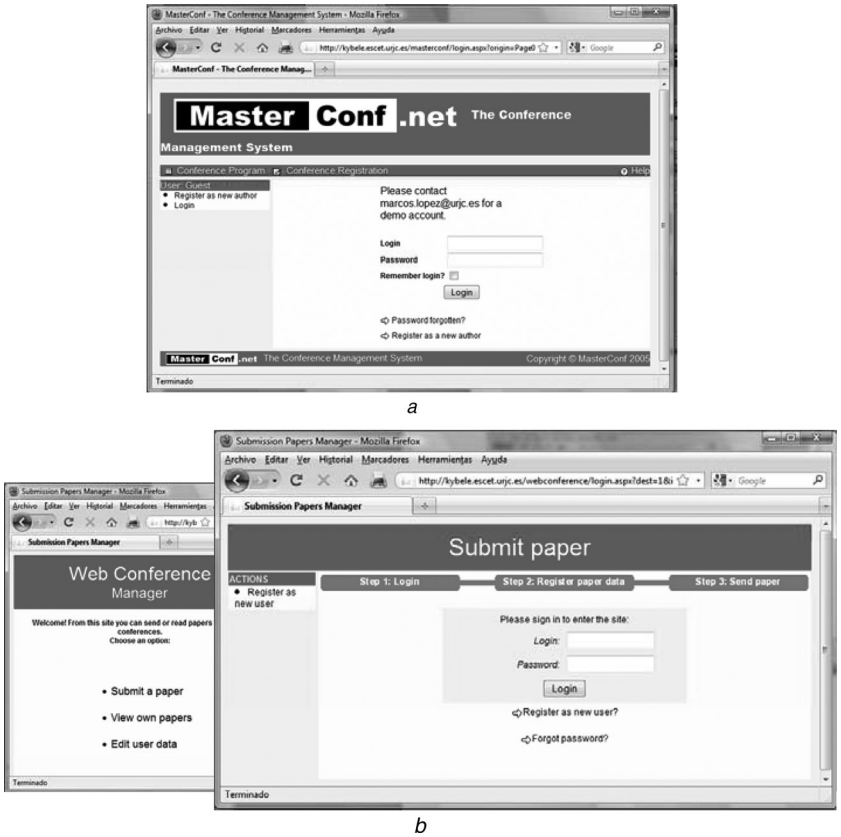


Fig. 2 *Extended navigation model of WebConference*



a

b

Fig. 3 *Screenshots*

a WebConference
b ConfMaster
For the purpose of this experiment we have developed a duplicate of the ConfMaster WIS denominated as MasterConf WIS

doi: 10.1049/iet-sen.2010.0062

used by the majority of web methodologies [7]: navigational classes, modelling pieces of information that will be shown on web pages and links between them, and the navigational structures such as menu, index, and so on. The method proposes the representation of two kinds of navigational classes, structural slices stereotyped by ⟪*SS*⟫ (representing data view) and functional slices stereotyped by ⟪*FS*⟫ (representing interaction with the user) [8]. This differentiation, however, is not necessary in other methodologies.

As a particular characteristic, the method proposes the design of the navigation model of the WIS based on two principal concepts: service and routes, the first of which is modelled as a use case in a use case diagram, and the second of which is modelled by means of an activity diagram. A more detailed description of the method can be found in [8]. However, in this work we present the design guidelines proposed to build the navigation in the WIS:

- The routes are explicitly identified in the extended navigation model. These routes are obtained from an activity diagram which models the process needed to carry out each service, basically representing a step of the route for each link of the activity diagram [30]. Each route is represented by an arrow stereotyped with ⟪*route*⟫ {*X*} where *X* is a description of the route. A route can be forked to provide alternative paths in the same service. These alternative paths are sub-routes of a route. A sub-route is represented by an arrow stereotyped with ⟪*route*⟫ {$X \cdot Y$}, where *Y* is a description of the sub-route of *X*.
- The extended navigation model also contains a main menu with the services needed by the user which represents the beginning of each route.

Although the navigation models of the two WISs are represented in this paper using the same notation, the design guidelines described previously have been applied only to the WebConference WIS. Thus, only that WIS has routes and the main menu indicating the beginning of the routes, which are the characteristics that we wish to study in this work. We shall now compare the two WISs, explaining the differences between them by means of their navigational model.

With regard to the ConfMaster, note that when an author accesses the WIS, s/he has three options (see main menu in Figs. 1 and 3*a*): 'login', 'register as a new author' and 'forgot password'. Nielsen affirms that the first immediate goal of any home page is to answer the questions: where am I? and what does this site do? [1]. In the ConfMaster WIS the main menu does not help users to answer these questions, since an author does not access the WIS to register as a new author, but in reality accesses it to submit a paper, to modify her/his data or to view her/his submitted paper. The extended navigation model of ConfMaster does not include routes.

Note that the WebConference WIS, unlike the ConfMaster WIS, presents a main menu with one entry for each service required by the user (see Figs. 2 and 3*b*).

Moreover, the extended navigation model of *WebConference* presents the routes through which to correctly carry out each service required by the user. So, for example, the route to submit a paper is represented by ⟪*route*⟫ {*SP*}, and ⟪*route*⟫ {*SP.RA*} represents the sub-route to register as a new author. These routes must also be taken into account during the implementation of the WIS. Fig. 3*b* shows how the route ⟪*route*⟫ {*SP*} is represented in the interface of the WIS.

With regard to navigation, Nielsen says that 'the navigational structure should be determined by the tasks users want to perform on your site' [1]. Thus, in WebConference, the entries in the main menu constitute the beginning of each route. So, for example, when a user wishes to submit a paper, the route will indicate that the author must first log in, that s/he must next register the paper's data and, finally, that s/he is able to send the paper.

The similarities and differences described above might lead to the belief that WebConference WIS has several characteristics that facilitate user navigation. Nevertheless, we believe that an empirical validation is needed to corroborate that a WIS with these design characteristics is more easily navigable that a WIS which does not include them.

4 Controlled experiment and its replica

This section provides a thorough explanation of the experimental process followed to run the experiment and its replica. The activities of the experimental process are based on the suggestions provided in [31, 32]. The experiment is reported by following the guidelines for reporting empirical research in software engineering [33] as closely as possible.

The experiment and its replica are described concurrently owing to the fact that their experimental phases are identical. A short communication of the initial experiment was presented in [34]. The principal differences between the original experiment and its replica are the way in which the links navigated during the experiment were collected, and the subjects' experience.

With regard to the first aspect, during the initial experiment all the experimental data were reported manually by the subjects. After the experiment we considered that a possible threat to internal validity might be the accuracy of the subjects' responses, given that they had to note down manually the time spent on doing the tasks and the names of the links navigated. In the replica this was therefore automatically recorded by computer logs, which considerably improved the data collection and, consequently, the accuracy of the results. All the links navigated were registered, and the time spent on each activity was exact.

With regard to the subjects' experience, the subjects who took part in the replica were considered to be more experienced, since they were PhD students and had been shown to have more experience in the debriefing questionnaire (task 1) that they had to fill out in the experiment execution (see Section 5).

More details of these differences are provided in the following sections in which the selected subjects and the data analysis are presented.

4.1 Goals and context

The main goal of this experiment and its replica, expressed by using the goal-question metric [35], is:

- To analyse the use of the proposed navigation model (which includes routes and a main menu indicating the beginning of each route) in the design of a WIS;
- For the purpose of characterisation;
- With regard to their support in users' navigability (*PerceivedEaseofUse, Efficiency, Effectiveness*);
- From the point of view of web engineering researchers;

• In the context of Computer Science undergraduate students from the Rey Juan Carlos University using two WISs concerned with conference management: WebConference (designed with routes and a main menu, including the beginning of each route) and ConfMaster (without routes and main menu, including the beginning of each route).

The subjects who participated in the family of experiments were:

• Experiment: 84 students enrolled in the 1st year of the Master's degree in Computer Science at the Rey Juan Carlos University (Spain).
• Replica: 16 PhD students in Information Technologies and Computing Systems at the Rey Juan Carlos University (Spain).

The subjects were students who had an average of 6 years of experience in using WISs and 3 years in designing them (see Table 3) and who were considered to be sufficiently competent to perform the level of experimental tasks required. However, they were not used to submitting papers to conferences. 'Convenience samples' were considered in both the experiment and its replica (i.e. all the students in all the available classes were selected). As was mentioned previously, the subjects of the replica were considered to be more experienced students.

As was also mentioned previously, the objects used in this experiment were the two WISs described in Section 3.

4.2 Independent and dependent variables and hypotheses formulation

When designing the experiment and its replica, it was necessary to consider what independent variables or factors were likely to make an impact on the results. For this experiment we considered one independent variable (main factor), called NM, which represents the use or non-use of the proposed navigation model (which includes routes and a main menu indicating the beginning of each route) in the design of a WIS. This variable has two treatments:

• *Use_NM:* This is represented by WebConference, which is a WIS designed using the navigation model with routes and a main menu indicating the beginning of each route.
• *Not_Use_NM:* This is represented by ConfMaster, which is a WIS designed without using the aforementioned navigation model.

In order to facilitate the reading of this paper, from here on we shall term the treatments as WebConference (Use_NM) and ConfMaster (Not_Use_NM).

We also considered other independent variables called 'co-factors'. These are:

• *Experience:* The subjects of the replica were considered to be more experienced since they are PhD students (Experience = 2), and the subjects of the experiment are Master students (Experience = 1). In addition, the subjects of the replica have an average of 2 years' more experience in using WIS (see Section 5). We therefore wished to investigate whether experience would have any influence on the results.
• *Order:* The subjects of the experiment were divided into two groups (G1, G2) (see Section 4.4.). In each group, the subjects received the experimental material in different order. We therefore wished to investigate whether the order in which the experimental material was received would affect the results.

We considered three dependent variables defined according to the ISO definition of usability (Ref ISO 9241-11) and [23]:

• *PerceivedEaseofUse*, defined as the degree to which a person believes that using a particular tool facilitates navigation through the WIS. This variable was measured through an ordinal measure obtained by means of Survey 1 included in Task 1 (see Section 4.3), which consists of eight questions rated using a five-point Likert scale.
• *Effectiveness*, defined as how well the use of a particular WIS allows the required tasks to be accomplished. This variable was measured as Number of Clicks/Number of Clicks the task required (a ratio scale measure). According to this definition, the closer to one the value is, the better the effectiveness. Section 4.3 shows details of Task 2, which was used to collect the data employed to calculate this measure.
• *Efficiency*, defined as the time spent in relation to how well the use of a particular WIS allows the required tasks to be accomplished. This variable was measured as the Number of Correct Clicks/Time (a ratio scale measure). According to this definition, the higher the value, the better the efficiency. Section 4.3 shows details of Task 2, which was used to collect the data employed to calculate this measure.

These variables measure the navigability of a WIS. In our case, navigability is the only sub-characteristic of the usability which varies in the two WIS that we are comparing, so we can infer that differences in these measures reflect the differences in the navigability of the two WISs.

We also wished to discover which WIS the subjects preferred to use, once they had used them. We therefore defined the fourth dependent variable:

• *Preference*, defined as the degree to which a person prefers a WIS once they have used it. This variable was measured through an ordinal measure obtained by means of Survey 2 included in Task 4 (see Section 4.3), which consists of eight questions rated using a three-point Likert scale.

The following main hypotheses were researched and shaped by our experience in WIS modelling:

• *Hypothesis 1:* $H_{0,1}$: There is no difference in the subjects' *PerceivedEaseofUse* when using WebConference and ConfMaster. $H_{1,1}$:$H_{0,1}$
• *Hypothesis 2:* $H_{0,2}$: There is no difference in preference between ConfMaster and WebConference. $H_{1,2}$:$H_{0,2}$
• *Hypothesis 3:* $H_{0,3}$: There is no difference in the subjects' *Efficiency* when using WebConference and ConfMaster. $H_{1,3}$:$H_{0,3}$
• *Hypothesis 4:* $H_{0,4}$: There is no difference in the subjects' *Effectiveness* when using WebConference and ConfMaster. $H_{1,4}$:$H_{0,4.}$

In order to test the effect of the co-factors (experience and order), we formulated the following hypothesis:

• *Hypothesis 5:* $H_{0,5}$: The subject's experience does not affect the *Effectiveness* when using a WIS. $H_{1,5}$:$H_{0,5}$.
• *Hypothesis 6:* $H_{0,6}$: The subject's experience does not affect the *Efficiency* when using a WIS. $H_{1,6}$:$H_{0,6}$.

doi: 10.1049/iet-sen.2010.0062

• *Hypothesis 7:* $H_{0,7}$: The order in which a subject uses a WIS does affect the *Effectiveness*. $H_{1,7}$:$H_{0,7}$.
• *Hypothesis 8:* $H_{0,8}$: The order in which a subject use a WIS does affect the *Efficiency*. $H_{1,8}$:$H_{0,8}$.

4.3 Experimental material and tasks

Each participant received the material to perform the following tasks:

• *Task 1:* This consisted of filling out a debriefing questionnaire, which included personal details and experience.
• *Task 2:* The subjects were required to use the corresponding WIS to perform the four tasks. In the experiment they were required to write down the data concerning the navigation (links, number of clicks and times) performed in the four tasks in the experiment. The manual registration of this data could have caused a certain amount of imprecision, which might have called into question the reliability of the data. We therefore decided to improve both WISs in the replica, permitting the automatic registration of these data through the computer logs. The subjects had to write down the IP address of the machine used in order to automatically register the navigation (links, number of clicks and times) performed by the students in the four tasks, through the computer logs. We had already performed these four tasks in advance, so we knew we had the correct links and the correct number of clicks needed to perform each task. All the links and clicks that were different from the correct ones were considered to be incorrect.
• *Task 3:* Filling in a survey consisting of eight questions related to the *PerceivedEaseofUse* of the WIS (Survey 1). For example, the first question (Q1) is: *Does the system provide the precise information needed to carry out the required tasks?*

1	2	3	4	5
not at all	very little	partially	quite a lot	yes, totally

• *Task 4:* On finishing the experimental tasks with both WISs, the subjects had to fill out Survey 2, which consisted of eight questions and in which they had to express their preference between the two WISs. The first question is, for example: *Which of the two WISs is better at providing the precise information needed to carry out the required tasks?*

1	2	3
ConfMaster	WebConference	Both the same

The list of required tasks and Surveys 1 and 2 are shown at the end of this paper, in Appendix. With regard to Survey 1, we present only that material which is related to the evaluation of WebConference, since the material for ConfMaster is analogous. The complete material is also available at: http://kybele.escet.urjc.es/nav_evaluation/material/.

4.4 Experimental design

The experiment and the replica were carried with two groups of subjects (G1 and G2), as is shown in Table 1. We selected a within subject design, that is, each subject was assigned the two treatments (see Table 2). The subjects of G1 first received the material related to ConfMaster and in then the material related to WebConference. The subjects of G2 received the material in a different order to cancel out learning effects; that is, they received the material for ConfMaster first and the material for WebConference second. The subjects were unaware of the hypothesis under investigation.

Table 1 Distribution of subjects per group

	Group 1 (G1)	Group 2 (G2)
experiment	44 subjects	40 subjects
replica	8 subjects	8 subjects

Table 2 Experimental design

	WebConference	ConfMaster
G1	First	Second
G2	Second	First

4.5 Experimental procedure

The experiment and the replica were executed in two sessions. In the first session the subjects received two hours of intensive training. In this session, they were given a test similar to those used in the experiment, and the tasks they had to carry out were explained to them. Once this sample test had been completed, we collected the data and checked whether the subjects had really understood the tasks.

In the second session, either the experiment itself or its replica was executed. Prior to that, the execution the subjects were assigned to each group randomly (see Table 1). The subjects worked under examination conditions, and were not permitted to talk to each other or to ask the professors supervising the experiment questions concerning any doubts they may have had. The subjects had to perform the tasks presented in Section 4.3.

We collected the material filled out by the subjects and ensured that it was complete. We discarded the data of two subjects in the experiment and one subject in the replica because they were incomplete. We originally had 86 subjects in the experiment, but only 84 of them are considered in the remainder of this paper in order to facilitate its reading. In the replica we originally we had 17 subjects, of which 16 are considered here, given that we discarded the data of one of them since it was incomplete.

Once the data had been collected, we calculated four measures of *Effectiveness* and *Efficiency* for each subject and object, one for each of the four tasks. In order to obtain only one *Effectiveness* (*Efficiency*) measure for each subject and object, we used the mean of these four measures.

The median of the measures obtained for *Preference* and *PerceivedEaseofUse* was similarly used in order to obtain one measure for each subject and each object.

4.6 Threats to validity

It is necessary to discuss the results of this experiment in order to determine any possible threats to validity. We focus on those threats presented in [32] that are most relevant to our experiment and replica. In our opinion, the greatest threats are to the internal validity of our experiment; that is, the degree to which conclusions can be drawn with regard to

the causal effect of the independent variable on the dependent variable [36]. One possible threat to internal validity is the accuracy of subject responses, given that they had to note down manually the time spent on doing the tasks and the names of the links navigated, although this aspect was improved in the replica by permitting the automatic registration of the links. Although special emphasis was placed on the relevance of the accuracy of these data during the training element of the experiment, complete certainty with regard to this accuracy was impossible and we had to rely on trust. We do not consider that subject motivation was a problem, since the students were motivated to participate in the experiment with a reward of 0.5 points for participating, and another 0.5 points for performing the required tasks correctly, in the final marks of the Database course.

Another threat to the internal validity is the difference between the number of subjects in the two levels of experience, which signifies that the design was unbalanced with regard to experience. However, this threat was decreased through the use of non-parametric tests [37].

With regard to external validity, that is, the ability to generalise the obtained findings to the population under study and other research settings [36], we consider that the functionality of the WISs selected was probably simple. The results therefore need to be confirmed by further experimentation.

Conclusion validity is concerned with the relationship between treatment and outcome, and the conclusions drawn from it. Two aspects must be considered: The first aspect concerns the appropriateness of the statistical tests. As reported above, we have screened our data for conformance with the assumptions of the statistical tests we used (Wilcoxon, Mann–Whitney). A Kolmogorov–Smirnov test confirmed that the normality assumption did not hold for the *Effectiveness*, *Efficiency*, *PerceivedEaseofUse* measures, and it was for this reason that non-parametric tests were selected to test the hypothesis. The second aspect concerns the effect sizes of the results. In the experiment we had 84 subjects, which can be considered as an appropriate sample size to solve potential issues regarding the statistical significance. The sample size in the replica is too small to consider the obtained results as being conclusive. The small sample size may have affected the results obtained.

With regard to construct validity, we shall now justify the appropriateness of the measures selected to measure the variables. In our case, the independent variable was measured with a nominal measure, NM, which takes two values (Use_NM, Not_Use_NM). We considered this measure to be appropriate for the purposes of the experiment. For the dependent variables, we selected two subjective ordinal scale measures, these being *PerceivedEaseofUse* and *Preference*. These measures were measured by using a Likert scale of 5 and 3 points, respectively. This type of measures is considered to be appropriate for measuring subjects' perceptions. We also considered objective ratio scale measures to measure the subjects' performance when carrying out the experiments, such as *Effectiveness* and *Efficiency*. These objective measures are commonly used to express the correctness of tasks and the relation between correctness and time spent, in experiments carried out in software engineering.

5 Data analysis and interpretation

The entire data analysis was carried out by means of SPSS [38]. We first summarised the data collected from the debriefing questionnaire in order to depict the subject profile (Table 3). The sample size was bigger in the original experiment, whereas in the replica the subjects had more experience and were older.

5.1 Analysis of the main factor

5.1.1 Testing the PerceivedEaseofUse hypothesis (H1): The first hypothesis was formulated to answer the question: 'Is WebConference (the object which was designed using the proposed navigation model) really easier to use than ConfMaster?'

The data used to test these hypotheses were the subjective ratings provided by the subjects in the first survey. We first checked the inter-rate reliability, to determine how consistent the results of the rates were with the order in which each WIS was used, by using Cronbach's alpha [39]. This coefficient is frequently used in IS-literature for this purpose (see, e.g. [40]). In the experiment, the Cronbach's alphas obtained for the responses were 0.81 for WebConference and 0.91 for ConfMaster, and in the replica they were 0.784 and 0.941, respectively. All these coefficient values were above 0.7, the suggested value to consider the results reliable. As these ratings were in an ordinal scale, we used the Wilcoxon test, a non-parametric test for comparing two independent samples with repeated measures, to test hypothesis 1. We tested this hypothesis for each question (Q1 to Q8) and also considered the median (Median Q1–Q8) of the eight responses, obtaining the results shown in Table 4(a) experiment and Table 4(b) (replica).

Table 3 Subjects' profile

	Min.	Max.	Mean	Standard deviation
(a) *Experiment*				
sex: 12 females – 72 males				
age	20	31	23.15	2.015
experience in using WIS	1	8	5.10	1.685
experience in designing WIS	0	5	1.51	1.393
experience in designing traditional IS	0	9	4.08	1.666
(b) *Replica*				
sex: 3 females – 13 males				
age	22	37	28.44	4.633
experience in using WIS	3	8	6.88	1.500
experience in designing WIS	1	7	3.81	1.721
experience in designing traditional IS	2	8	5.13	1.668

doi: 10.1049/iet-sen.2010.0062

Table 4 Wilcoxon test

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Median Q1–Q8
(a) *Experiment*									
Z	−2.472	−3.116	−2.055	−3.088	−4.084	−3.139	−2.288	−2.586	−4.076
significance level	0.013	0.002	0.040	0.002	0.000	0.002	0.022	0.010	0.000
(b) *Replica*									
Z	−1.732	−2.723	−1	−1.732	−1.603	−1.997	−2.233	1.603	−1.603
significance level	0.083	0.006	0.317	0.083	0.108	0.045	0.025	0.108	0.108

The experimental results presented in Table 4(a) show that as all the significance levels are lower than 0.05 we can reject $H_{0.1}$. This means that the *PerceivedEaseofUse* was different in ConfMaster and WebConference in the experiment. In the replica the significance levels are lower than 0.05 only for questions Q2 and Q6 (Table 4(b)). We cannot, therefore, reject the null hypothesis for all the questions, or for the median of all values (Median Q1–Q8 in Table 4(b)).

However, upon analysing the Box-plots shown in Figs. 4 and 5, which compare the median (Median Q1–Q8) of the responses obtained for each WIS, we can conclude that the *PerceivedEaseofUse* is better for the WebConference, both in the experiment and in its replica.

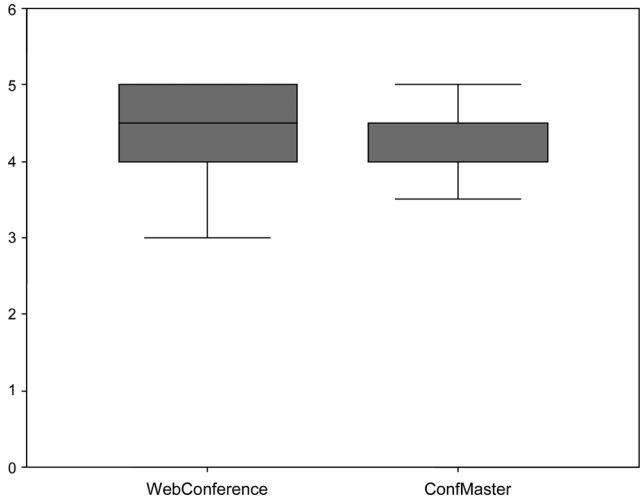


Fig. 4 *Box-plot comparing the median of PerceivedEaseofUse of WebConference and ConfMaster (experiment)*

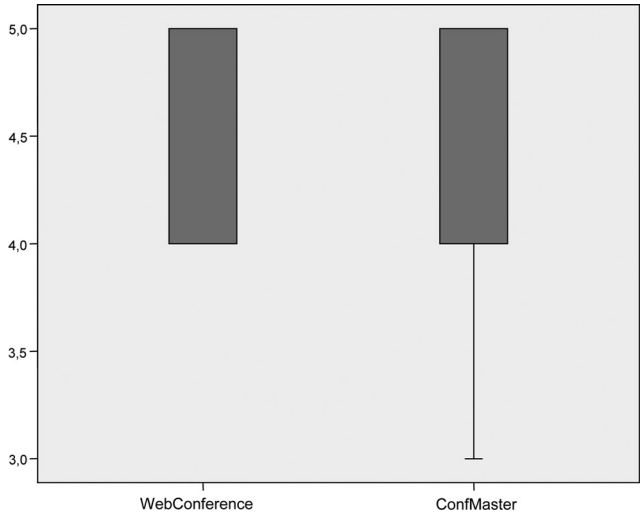


Fig. 5 *Box-plot comparing the median of PerceivedEaseofUse of WebConference and ConfMaster (replica)*

It is interesting to highlight that it was the subjects with less experience (experiment) whose *PerceivedEaseofUse* was significantly different when comparing ConfMaster with WebConference.

Of the subjects who participated in the experiment 67% rated WebConference with a value (Median Q1–Q8) greater than 4. This demonstrates that the majority of the subjects perceived WebConference to be easier to use.

A similar situation arises in the replica, given that 53% of the subjects rated WebConference with a value (Median Q1–Q8) greater than 4.

5.1.2 Testing the preference hypothesis (H2): The second hypothesis was formulated to answer the question: 'Which WIS do the subjects prefer to use, WebConference (the object that was designed by using the proposed navigation model) or ConfMaster?' To test this hypothesis we used the data obtained in Survey 2, which was assigned to the subjects after they had performed the experimental tasks with both WISs.

As a result, we found that 65 subjects preferred WebConference, 15 preferred ConfMaster and only 4 did not have a preference. In the replica, 11 subjects preferred WebConference, 2 preferred ConfMaster and only 3 did not have a preference (see Table 5). By analysing the results of the test of the comparison of proportions shown in this table, we obtained a P-value <0.001, and the estimation of the proportions were greater for WebConference (above 0.6 in both cases) which suggests that we should reject $H_{2,0}$, thus confirming that there is a greater probability that the subjects prefer WebConference to ConfMaster. The results also reveal that ConfMaster is less preferred by the subjects in the replica, who have more experience.

5.1.3 Testing the Effectiveness and Efficiency hypothesis (H3 and H4): Before testing *H*3 and *H*4, we shall present the descriptive statistics of the *Effectiveness*

Table 5 Comparison of proportions regarding preferences between WebConference and ConfMaster

	Experiment	Replica
WebConference	65	11
ConfMaster	15	2
no preference	4	3
$P_{WebConference}$	0.6415	0.6875
$P_{ConfMaster}$	0.2830	0.1250
P-value	0.000	0.000

Table 6 Descriptive statistics of effectiveness and efficiency

Statistics	Effectiveness		Efficiency	
	WebConference	ConfMaster	WebConference	ConfMaster
(a) *Experiment*				
size	84	84	80	42
ranging from	0 to 5.75	0 to 3.2	0,.011 to 0.151	0.008 to 0,092
mean	1.532	0,568	0.046	0.031
median	1.365	0.125	0.038	0.024
standard deviation	0.927	0.672	0.026	0.018
(b) *Replica*				
size	16	16	16	16
ranging from	0.917 to 1.33	0.688 to 1.229	0.037 to 0.282	0.030 to 0.154
mean	1.047	1.047	0.091	0.078
median	1.000	1.021	0.078	0.073
standard deviation	0.130	0,138	0.058	0.037

and *Efficiency* in the experiment and in the replica, in order to obtain a global description of them. Table 6(a) presents the results of the experiment, which indicate that the median and the mean of *Effectiveness* are closer to 1 when the subjects use WebConfernce and, furthermore, that the standard deviation is better. The median and the mean of the *Efficiency* are also better when the subjects use WebConference. Table 6(b) shows the results of the replica, which were similar to those of the experiment. This signifies that the navigability is better when WebConference is used, independently of whether the subjects have more or less experience.

The normality of the data was analysed by using the Kolmogorov–Smirnov test. Since we found that the data were not normal, we decided to use the Wilcoxon signed rank test, which is a non-parametric test for paired samples. We studied the medians of *Effectiveness* (*Efficiency*) in the experiment and in the replica. In the experiment the results were significant (P-value < 0.001) for both variables, so we can accept that the use of WebConference in the experiment improves navigability.

In the replica, the *Effectiveness* results were not significant (P-value ≥ 0.5), and neither were those of *Efficiency* (P-value > 0.10). This may signify that the effect of using WebConference is different depending on the subjects' experience; that is, those subjects with less experience take more advantage of the WIS with routes and the beginning of each route. We consider that this is a logical result owing to the fact that these characteristics help the users navigate through the WIS, and subjects with less experience might need this help. However, more experienced subjects usually prefer more freedom when navigating, and it may be that the routes do not assist them to carry out a task to the same extent. Nonetheless, the size of the replica is not particularly large, and the test power is low (<0.5), so the results obtained in the replica cannot be considered as conclusive.

5.2 Analysis of experience (H5, H6)

In this section we shall analyse whether there is any difference in terms of *Effectiveness* and *Efficiency* between the subjects who took part in the experiment and those who participated in the replica. Remember that we considered the subjects in the replica to be more experienced. Before testing *H*5 and *H*6, we studied the descriptive statistics obtained. Table 7 shows that the median and the mean of *Effectiveness* are closer to 1 when the subjects have more experience, and the standard deviation is also better. The median and the mean of *Efficiency* are better in the replica, that is, for the more experienced subjects.

We decided to use the Mann–Whitney test, which is a non-parametric test for comparing the medians obtained for the *Effectiveness*/*Efficiency* in the experiment and in the replica (i.e. independent samples), because the data were not normally distributed. The P-value is 0.191 for *Effectiveness* and 0.000 for *Efficiency*. In conclusion, there is no evidence to allow us to accept that experience affects *Effectiveness*. However, experience does appear to affect *Efficiency*. This signifies that more experienced subjects obtain more correct results in less time.

5.3 Analysis of order (H7, H8)

We decided to use the Mann–Whitney test for testing *H*7 and *H*8 in both the experiment and the replica (i.e. independent samples), because the data were not normally distributed.

Table 7 Descriptive statistics of effectiveness and efficiency

Statistics	Experiment		Replica	
	Effectiveness	Efficiency	Effectiveness	Efficiency
size	84	84	16	16
ranging from	0 to 2.875	0.0105 to 0.151	0.844 to 1.197	0.044 to 0.218
mean	1.050	0.0426	1.047	0.0842
median	0.915	0.037	1.047	0.074
standard deviation	0.520	0.025	0.087	0.042

doi: 10.1049/iet-sen.2010.0062

The test for WebConference compares two medians: the median of the *Efficiency_WebConference* (*Effectiveness_ WebConference*) observed when WebConference is used first (G1) and the median when it is used second (G2), in the experiment and in the replica. The test for ConfMaster compares two medians: the median of the *Efficiency_ ConfMaster* (*Effectiveness_ConfMaster*) observed when the ConfMaster is used first (G2) and the median when it is used second (G1), in the experiment and in the replica.

The results were obtained using the Mann–Whitney test (see Table 8). When analysing WebConference, we cannot reject *H*7,0 and *H*8,0, given that all the *P*-values > 0.05. This signifies that the order does not influence the use of WebConference. The results of ConfMaster, in which all the *P*-values < 0.05, reveal that the order affects the use of ConfMaster, rejecting thus *H*7,0 and *H*8,0. This to some extent confirms that the improvement in *Effectiveness*/*Efficiency* that is obtained when WebConference is used results solely from the proposed navigation model used in its design, and not from the order in which the WIS were used. This effect is, moreover, the same for subjects with a greater or a lesser degree of experience.

Table 8 *P*-values of Mann–Whitney test (*H*7, *H*8)

P-values	WebConference		ConfMaster	
	Effectiveness	Efficiency	Effectiveness	Efficiency
Experiment	0.567	0.874	0.04	0.06
Replica	0.101	0.958	0.003	0.007

5.4 Summary of findings and discussion

Table 9 presents the summary of the finding obtained from the data analysis. Upon analysing Table 9, we obtained the following findings:

- *Hypothesis 1:* In the experiment, the *PerceivedEaseofUse* is different in ConfMaster and WebConference. In fact, the ratings when using WebConference are higher than those used in ConfMaster. It would appear that the use of the proposed navigation model (which includes routes and a main menu indicating the beginning of each route) improves the subjects' perception as regards the navigability of a WIS. However, the significance levels in the replica are lower than 0.05 only for questions Q2 and Q6, signifying that it is not clear whether with the more experienced group the use of the proposed navigation model improves the perception of navigability. Nevertheless, the medians were higher WebConference in all the questions, and significant differences were perceived solely in questions Q2 and Q6. 67% of the subjects who participated in the experiment rated WebConference with a value (the median of the 8 the ratings of the 8 questions (Median Q1–Q8)) greater than 4. This demonstrates that the majority of the subjects perceived WebConference to be easier to use. A similar situation arises in the replica, given that 53% of the subjects

Table 9 Summary of the hypothesis testing (S = significant; NS = no significant)

Hypothesis	Results
Hypothesis 1: $H_{0,1}$: There is no difference in subjects' *PerceivedEaseofUse* when using WebConference and ConfMaster. $H_{1,1}$: $H_{0,1}$	*Experiment* PerceivedEaseofUse_WebConference > PerceivedEaseofUse_ConfMaster(S) *Replica* PerceivedEaseofUse_WebConference > PerceivedEaseofUse_ConfMaster(NS)
Hypothesis 2: $H_{0,2}$: There is no difference in *Preference* between ConfMaster and WebConference. $H_{1,2}$: $H_{0,2}$	*Experiment* Preference_WebConference > Preference_ConfMaster(S) *Replica* Preference_WebConference > Preference_ConfMaster(S)
Hypothesis 3: $H_{0,3}$: There is no difference in subjects' *Efficiency* when using WebConference and ConfMaster. $H_{1,3}$: $H_{0,3}$	*Experiment* Effectiveness_WebConference < Effectiveness_Confmaster(S) Effiicciency_WebConference < Efficiency_ConfMaster(S)
Hypothesis 4: $H_{0,4}$: There is no difference in subjects' *Effectiveness* when using WebConference and ConfMaster. $H_{1,4}$: $H_{0,4}$	*Replica* Effectiveness_WebConference < Effectiveness_ConfMaster(NS) Effiicciency_WebConference < Efficiency_ConfMaster(NS)
Hypothesis 5: $H_{0,5}$: Subjects' experience does not affect *Effectiveness* when using a WIS. $H_{1,5}$: $H_{0,5}$	Effectiveness(NS)
Hypothesis 6: $H_{0,6}$: Subjects' experience does not affect the *Efficiency* when using a WIS. $H_{1,6}$: $H_{0,6}$	Efficiency(S)
Hypothesis 7: $H_{0,7}$: The order in which a subject uses a WIS does affect the *Effectiveness.* $H_{1,7}$: $H_{0,7}$	*Experiment* Effectiveness_WebConference_G1 = Effectiveness_ WebConference_G2 (NS)
Hypothesis 8: $H_{0,8}$: The order in which a subject uses a WIS does affect the *Efficiency.* $H_{1,8}$: $H_{0,8}$	Efficiency_WebConference_G1 = Efficiency_WebConference_G2(NS) Effectiveness_ConfMaster_G1= Effectiveness_ConfMaster_G2(S) Efficiency_ConfMaster_G1 = Efficiency_ConfMaster_G2(S) *Replica* Effectiveness_WebConference_G1 = Effectiveness_WebConference_G2(NS) Efficiency_WebConference_G1 = Efficiency_WebCoference_G2(NS) Effectiveness_ConfMaster_G1 = Effectiveness_ConfMaster_G2(S) Efficiency_ConfMaster_G1 = Efficiency_ConfMaster_G2(S)

rated WebConference with a value (Median Q1–Q8) greater than 4.

- *Hypothesis 2:* The results confirm that there is a greater probability that the subjects prefer WebConference. Moreover, ConfMaster is less preferred by the subjects with more experience (replica). This shows that, independent of their level of experience, the subjects prefer using the WIS designed using the proposed navigation model.
- *Hypothesis 3, 4:* The use of the navigation model proposed improves the navigability of WISs (measured by *Effectiveness* and *Efficiency*) in the experiment, that is, with less experienced subjects. However, these results are not conclusive in the replica owing to the fact that the sample size was small.
- *Hypothesis 5, 6:* The *Efficiency* is affected by experience, that is, the subjects with more experience (replica) make more correct clicks per second than the subjects with less experience (experiment). This means that more experienced subjects obtain more correct results in less time. The *Effectiveness* is not affected by experience in the case of those subjects with less experience (experiment), but this insignificant result is not conclusive, since the sample size of the replica was small.
- *Hypothesis 7, 8:* The order in which a WIS designed using the proposed navigation model (which includes routes and a main menu indicating the beginning of each route) is used does not affect navigability (measured by *Effectiveness* and *Efficiency*). However, the order has an influence when a WIS designed without using the proposed navigation model is used. This, to some extent, confirms that the improvement in *Effectiveness*/*Efficiency* obtained when using the WIS built with the proposed navigation model is solely owing to its use in the design of the proposed navigation model and not to the order in which this WIS is used. This effect is, moreover, the same for all subjects, regardless of their level of experience.

These findings reveal that the use of routes and a main menu, including the beginning of each route, improves the navigability of the WIS built from it, especially for subjects with less experience. Moreover, the subjective perceptions of the subjects with less experience (measured by using *PerceivedEaseofUse* and *Preference*) coincide with the objective measures for navigability (*Effectiveness* and *Efficiency*). For more experienced subjects the results are not conclusive given the small sample size. This to some extent demonstrates that the subjects with less experience take more advantage of the routes because they only need to navigate by following the routes to carry out a task, and in the other model in which there are no routes it is easier to get lost.

6 Implications

The principal conclusion obtained from the empirical study is that, given two WISs with the same functionality and identical interface styles, the users prefer the WIS built using the proposed navigation model, which includes routes and a main menu including the beginning of each route, and that they also perceive it to be more easily navigable, as we have empirically corroborated.

However, the most important aspect of our findings is that, starting from a navigation model with certain particular characteristics, it is possible to build a more easily navigable WIS. We obviously agree that an experienced web designer could obtain a system with similar characteristics to those analysed in this paper, that is, with a main menu including the beginning of each service provided by the application and routes to signpost the steps to carry each service out. Nevertheless, the importance of our study focuses on identifying certain design characteristics which, when present in a navigation model, allow more easily navigable WISs to be systematically built. A good design is not therefore dependent on the web designer's experience or skill.

The navigability of WISs is expressed in web engineering methods through navigational models [7]. However, most of the main web engineering methods [41–46] focus solely on the analysis and design of the navigation [7], leaving the final implementation, and consequently some aspects that affect the WIS' usability (such as that tested in this work), to the web designer's discretion.

This aspect is a particularly crucial in determining success factor in certain types of web applications. Deshpande *et al.* presented taxonomy of WISs in [47], which is summarised in Table 10. From our point of view, design characteristics such as those analysed in this work may be of particular importance in the presence of high behaviour content, that is, applications in which more or less complex tasks or processes are carried out, such as transaction, workflow, collaborative work environments, online communities, web portals or web services. The design aspect of routes and main menus including the beginning of each route may be of less importance in applications which are more data-oriented or informational, such as those which are designed to show information and carry out simple modification or querying tasks.

7 Conclusions and future work

In this work we have presented an experiment and its replica to test whether the use of routes and a main menu including the beginning of each route effectively makes it possible to

Table 10 Categories of WISs [47]

Category	Examples
informational	online newspapers, product catalogues, newsletters, service manuals, classifieds, e-books
interactive	registration forms, customised information presentation, games
transaction	e-shopping, ordering goods and services, banking
workflow	planning and scheduling systems, inventory management, status monitoring
collaborative work environments	distributed authoring systems, collaborative design tools
online communities, marketplaces	chat groups, recommender systems, marketplaces, auctions
web portals	electronic shopping malls, intermediaries
web services	enterprise applications, information and business intermediaries

doi: 10.1049/iet-sen.2010.0062

obtain greater navigability, and to build more easily navigable WISs. The experiment was carried out with less experienced subjects (84 subjects), while the subjects of the replica were more experienced (16 subjects). The most important conclusions obtained from the empirical study are that subjects, especially those with less experience (1st year Master students of Computer Science with in average 5 years using WISs and 1.5 years designing them), perceive WebConference (the WIS that was built using routes and a main menu including the beginning of each route) to be easier to use, since they used it more effectively, that is, the use of WebConference led them to perform the required tasks more correctly. In addition, they used them more efficiently, that is, they produced more correct results in less time. The results obtained for more experienced subjects reveal a slight tendency in favour of the WIS built using the proposed navigation model, although these results are not conclusive owing to the small sample size.

Although the results obtained are encouraging, we consider that further validation is needed to obtain conclusive results with regard to whether the proposed design guidelines really led to the creation of WISs which are more effective, more efficient and easier to use. We are therefore planning to carry out new replications of this experiment with bigger samples size and with more complex WISs, in which the results should be more conclusive, since the WISs evaluated in this paper are quite simple.

The design characteristics used to build WISs in a methodological manner, including routes and the beginning of the routes, were proposed in the framework of a hypertext modelling method, although they can be applied to various navigation modelling methods. The design principles focus on modelling the navigation of WISs from a user need oriented perspective. They thus permit a navigation model to be built with various characteristics that could facilitate a user's navigation through WISs, such as a main menu with the services required by the user and the signposting of the sequence of steps with which to correctly carry out each service, called routes. The principal benefit of our research is the identification of certain design characteristics which will allow more easily navigable WISs to be built in a systematic manner, independently of web designers' skills.

8 Acknowledgments

This research has been funded by the following projects: MODEL CAOS (MICINN TIN2008-0358), MEDUSAS (CDTI-MICINN and FEDER IDI-20090557), ORIGIN (CDTI-MICINN and FEDER IDI-2010043 (1-5)), PEGASO/MAGO (MICINN and FEDER, TIN2009-13718-C02-01), EECCOO (MICINN TRA2009_0074) MECCA (JCMM PII2I09-0075-8394) and IMPACTUM (PEII 11-0330-4414). We would also like to thank Marcos López for implementing the WebConference WIS and the students who participated in the experiments.

9 References

1 Nielsen, J.: 'Designing web usability' (New Riders Publishing, 2000)
2 El-Ramly, M., Stroulia, E.: 'Analysis of web-usage behavior for focused web sites: a case study', *J. Softw. Maintenance Evol.: Res. Pract. Archive*, 2004, **16**, (1–2), pp. 129–150
3 Constantine, L.L., Lockwood, L.A.D.: 'Software for use: a practical guide to the models and methods of usage centered design' (Addison-Wesley, 1999)
4 Mendes, E., Moesly, N. (Eds.): 'Web engineering' (Springer, 2005)
5 Palmer, J.: 'Designing for web site usability', *IEEE Comput.*, 2002, **35**, (7), pp. 102–103
6 Palmer, J.: 'Web site usability, design and performance metrics', *Inf. Syst. Res.*, 2002, **13**, (2), pp. 151–167
7 Escalona, M.J., Torres, J., Mejías, M., Gutiérrez, J.J., Villadiego, D.: 'The treatment of navigation in web engineering', *Adv. Eng. Softw.*, 2007, **38**, (4), pp. 267–282
8 De Castro, V., Marcos, E., Cáceres, P.: 'A user service oriented method to model web information systems'. Proc. Int. Conf. on Web Information Systems Engineering (WISE'04), 2004, (*LNCS*, **3306**), pp. 41–52
9 Zelkowitz, M., Wallace, D.: 'Experimental validation in software engineering', *Inf. Softw. Technol.*, 1997, **39**, (11), pp. 735–743
10 Basili, V., Shull, F., Lanubile, F.: 'Building knowledge through families of experiments', *IEEE Trans. Softw. Eng.*, 1999, **25**, (4), pp. 456–473
11 Conte, T., Massollar, J., Mendes, E., Travassos, G.: 'Usability evaluation based on web design perspectives'. Proc. First Int. Symp. on Empirical Software Engineering and Measurement (ESEM 2007), Madrid, Spain, 2007, pp. 146–155
12 ConfMaster: http://www.confmaster.net/, accessed 2011
13 WebConference: http://kybele.escet.urjc.es/webconference-app/, accessed 2011
14 Cheung, C.M.K., Lee, M.K.O.: 'The structure of web-based information systems satisfaction: testing of competing models', *J. Am. Soc. Inf. Sci. Technol.*, 2008, **59**, (10), pp. 1617–1630
15 Bruno, V., Tam, A., Thom, J.: 'Characteristics of web applications that affect usability: a review'. ACM Int. Conf. Proc. 17th Australia Conf. on Computer–Human Interaction (HCI'05), 2005, pp. 1–4
16 Zhang, P., Von Dran, G.M.: 'Satisfiers and dissatisfiers: a two-factor model for website design and evaluation', *J. Am. Soc. Inf. Sci.*, 2000, **51**, (14), pp. 1253–1268
17 Calongne, C.: 'Designing for web site usability'. Proc. Seventh Annual Consortium for Computing in Small Colleges Central Plains Conf. on Journal of Computing in Small Colleges, 2001, pp. 39–45
18 Shneiderman, B.: 'Designing the user interface: strategies for effective human–computer interaction' (Addison-Wesley, 1998)
19 Yu, B.M., Roh, S.Z.: 'The effects of menu design on information-seeking performance and user's attitude on the world wide web', *J. Am. Soc. Inf. Sci. Technol.*, 2002, **53**, (11), pp. 923–933
20 Juvina, I., Van Oostendorp, H.: 'Predicting user preferences: from semantic to pragmatic metrics of web navigation behavior'. ACM Proc. Conf. on Dutch Directions in HCI (HCI'04), 2004, pp. 1–10
21 Gwizdka, J., Spence, I.: 'Predicting outcomes of web navigation'. Special Interest Tracks and Posters of the 14th Int. Conf. on World Wide Web (WWW '05), 2005, pp. 333–336
22 Burrell, A., Sodan, A.: 'Web interface navigation design: which style of navigation-link menus do users prefer?'. IEEE Proc. 22nd Int. Conf. on Data Engineering Workshops (ICDEW'06), 2006, pp. 42–51
23 Davis, F.D.: 'Perceived usefulness, perceived ease of use and user acceptance of information technology', *MIS Quart.*, 1989, **3**, (3), pp. 318–341
24 Matera, M., Costabile, M.F., Garzotto, F., Paolini, P.: 'SUE inspection: an effective method for systematic usability evaluation of hypermedia', *IEEE Trans. Syst. Man, Cybern.*, 2002, **32**, pp. 93–103
25 Dix, A., Finlay, J., Abowd, G., Beale, R.: 'Human–computer interaction' (Prentice-Hall, Englewood Cliffs, NJ, 1998, 2nd edn.)
26 Nielsen, J., Mack, R.L.: 'Usability inspection methods' (Wiley, New York, 1994)
27 Polson, P.G., Lewis, C., Riemanand, J., Wharton, C.: 'Cognitive walkthroughs: a method for theory-based evaluation of user interfaces', *Int. J. Man–Mach. Stud.*, 1993, **36**, pp. 741–773
28 Kock, E., van Biljon, J., Pretorius, M.: 'Usability evaluation methods: mind the gaps'. Proc. 2009 Annual Research Conf. South African Institute of Computer Scientists and Information Technologists, South Africa, 2009, pp. 122–131
29 Koch, N.: 'Software engineering for adaptive hypermedia systems: reference model, modeling techniques and development process'. PhD thesis, Ludwig-Maximilians-Universität München, 2002
30 Cáceres, P., De Castro, V., Vara, J.M., Marcos, E.: 'Model transformations for hypertext modeling on web information systems'. Proc. 2006 ACM Symp. on Applied Computing (SAC), Dijon, France, 23–27 April 2006, pp. 1232–1239

31 Ciolkowski, M., Shull, F., Biffl, S.: 'A family of experiments to investigate the influence of context on the effect of inspection techniques'. Proc. Int. Conf. on Empirical Assessment in Software Engineering, 2002, pp. 48–60
32 Wohlin, C., Runeson, P., Host, M., Ohlsson, M.C., Regnell, B., Wesslen, A.: 'Experimentation in software engineering: an introduction' (Kluwer Academic Publishers, 2000)
33 Jedlitschka, A., Ciolkowoski, M., Pfahl, D.: 'Reporting experiments in software engineering', in Shull, F., Singer, J., Sjøberg, D. (Eds.): 'Guide to advanced empirical software engineering' (Springer, 2008)
34 De Castro, V., Genero, M., Marcos, E., Piattini, M.: 'Including routes in web information systems as a way to improve the navigability: an empirical study'. Proc. Web Information Systems Engineering Workshops, 2007, (*LNCS*, **4832**), pp. 505–510
35 Basili, V., Rombach, H.: 'The TAME project: towards improvement-oriented software environments', *IEEE Trans. Softw. Eng.*, 1998, **14**, (6), pp. 758–773
36 Briand, L., Bunse, C., Daly, J.: 'A controlled experiment for evaluating quality guidelines on the maintainability of object-oriented designs', *IEEE Trans. Softw. Eng.*, 2001, **27**, (6), pp. 513–530
37 Siegel, S.: 'Estadística no paramétrica' (Trillas, México, 1985)
38 SPSS 11.5: 'Syntax reference guide' (SPSS Inc., Chicago, 2002)
39 Cronbach, L.J.: 'Coefficient alpha and the internal structure of tests', *Psychometrika*, 1951, **16**, (3), pp. 297–334
40 Shanks, G.: 'Conceptual data modeling: an empirical study of expert and novice data modelers', *Australian J. Inf. Syst.*, 1997, **4**, (2), pp. 63–73
41 Fraternali, P.: 'Tools and approaches for developing data-intensive web applications: a survey', *ACM Comput. Surveys*, 1999, **31**, (3), pp. 227–263
42 Garzotto, F., Paolini, P., Schwabe, D.: 'HDM – a model-based approach to hypertext application design', *ACM Trans. Database Syst.*, 1993, **11**, (1), pp. 1–26
43 Gómez, J., Cachero, C., Pastor, O.: 'Conceptual modeling of device-independent web applications', *IEEE Multim.*, 2001, **8**, (2), pp. 26–39
44 Hennicker, R., Koch, N.: 'A UML-based methodology for hypermedia design'. Proc. UML Conf. (UML'00), (*LNCS*, **1939**), 2000, pp. 410–424
45 Isakowitz, T., Kamis, A., Koufaris, M.: 'The extended RMM methodology for web publishing'. Working Paper IS-98-18, Center for Research on Information System, 1998. Retrieved from http://rmm-java.stern.nyu.edu/rmm/
46 Knapp, A., Koch, N.: 'Model-driven generation of web applications in UWE'. Proc. Model Driven Web Engineering, CEUR-WS, 2007, vol. 261
47 Deshpande, Y., Murugesan, S., Ginige, A., *et al.*: 'Web engineering', *J. Web Eng.*, 2002, **1**, (1), pp. 3–17
48 Dybå, T., Kitchenham, B.A., Jorgensen, M.: 'Evidence-based software engineering for practitioners', *IEEE Softw.*, 2005, **22**, (1), pp. 58–65
49 Miller, J.: 'Applying meta-analytical procedures to software engineering experiments', *J. Syst. Softw.*, 2000, **54**, (1), pp. 29–39

Q4

10 Appendix: experimental material

This appendix includes part of the experimental material. The complete material can be found at: http://kybele.escet.urjc.es/nav_evaluation/material/.

10.1 Questionnaire regarding personal details (Task 1)

Personal data

Sex: Male ☐ Female ☐

Age: _____

Course: _____

Years of experience using web applications: _____

Years of experience in web application design: _____

Years of experience in design of traditional applications: (Not Web): _____

Name and Surname (Optional): _____

10.2 List of tasks to be performed (Task 2)

Task 1: Send a paper

You should send from de: http://kybele.escet.urjc.es/Documentos/exper/index.htm

PLEASE FILL IN ONLY THAT DATA WHICH IS REQUESTED (marked with an ∗).

Put your own name as that of the first author.

Have you completed this activity correctly? If not, please state why:

Task 2: Modify the city in your address.

Task 3: Send new paper.

You must send from de: http://kybele.escet.urjc.es/Documentos/exper/index.htm

PLEASE FILL IN ONLY THAT DATA WHICH IS REQUESTED (marked with an ∗)

Put your own name as that of the first author.

Have you completed this activity correctly? If not, please state why:

Task 4: Look at the papers which have been sent and download one of them.

10.3 Survey 1: including eight questions related to the PerceivedEaseofUse of the WIS (Task 3)

Evaluation questionnaire

Please read the following questions carefully before answering.

Put a circle around the number which corresponds with your reply.

1. Does the system provide the precise information needed to carry out the solicited tasks?

1	2	3	4	5
not at all	very little	partially	quite a lot	yes, totally

2. Do you consider that the system is easy to use?

1	2	3	4	5
not at all	very little	partially	quite a lot	yes, totally

3. If you were to organise a conference, would you use this application to send the papers?

1	2	3	4	5
not at all	very little	partially	quite a lot	yes, totally

4. Would it be easy for you to become an expert in the use of this tool?

1	2	3	4	5
not at all	very little	partially	quite a lot	yes, totally

5. When carrying out a task, do you know exactly what the next step that must be taken is?

1	2	3	4	5
not at all	very little	partially	quite a lot	yes, totally

6. Is it easy to change from one task to another without getting lost?

1	2	3	4	5
not at all	very little	partially	quite a lot	yes, totally

doi: 10.1049/iet-sen.2010.0062

7. Was it easy to find the correct link with which to carry out the solicited tasks?

1	2	3	4	5
not at all	very little	partially	quite a lot	yes, totally

8. Did you feel comfortable using this application?

1	2	3	4	5
not at all	very little	partially	quite a lot	yes, totally

10.4 Survey 2: including questions to compare the two WISs (Task 4)

Tools comparison questionnaire

Please read the following questions carefully before answering:

Put a circle around the number which corresponds with your reply.

1. Which of the two systems is better at providing the precise information needed to carry out the tasks solicited?

1	2	3
MasterConf	WebConference	Both the same

2. Which of the two systems do you consider to be easiest to use?

1	2	3
MasterConf	WebConference	Both the same

3. If you were to organise a conference, which of the two applications would you use to send the papers?

1	2	3
MasterConf	WebConference	Both the same

4. With which of the two tools would it be easiest to become an expert in its use?

1	2	3
MasterConf	WebConference	Both the same

5 In which of the two systems is it easiest to know exactly what the next step that must be taken to complete a task is?

1	2	3
MasterConf	WebConference	Both the same

6. In which of the two systems is it easiest to change from one task to another without getting lost?

1	2	3
MasterConf	WebConference	Both the same

7. In which of the two systems is it easiest to find the correct link to carry out the solicited tasks?

1	2	3
MasterConf	WebConference	Both the same

8. Which of the two systems was most comfortable for you to use?

1	2	3
MasterConf	WebConference	Both the same

www.ietdl.org

SEN20100062

Author Queries

V. de Castro, M. Genero, E. Marcos, M. Piattni

Q1 Please reduce the size of the Abstract so that it does not exceed 200 words.

Q2 Please note that the references are renumbered to be in numerical order as per the journal style.

Q3 Please check the initial citation of Table 3 in text.

Q4 References [48, 49] are not cited in text. Please cite it in the text or delete it from the reference list.