PREPRINT VERSION

Evaluating Interface Design Choices on WAP Phones: Navigation and Selection

LUCA CHITTARO AND PAOLO DAL CIN

HCI Lab, Dept. of Math and Computer Science
University of Udine
via delle Scienze 206, 33100 Udine, ITALY
+39 0432 558450

The published article appeared in:

Personal and Ubiquitous Computing, vol. 6, no. 4, 2002, pp. 237-244.

Evaluating Interface Design Choices on WAP Phones: Navigation and Selection

LUCA CHITTARO AND PAOLO DAL CIN

HCI Lab, Dept. of Math and Computer Science
University of Udine
via delle Scienze 206, 33100 Udine, ITALY
+39 0432 558450
chittaro@dimi.uniud.it

Abstract. Wireless Application Protocol (WAP) phones are a growing relevant part of the mobile market, and the number of offered WAP services is rapidly increasing. Usability is crucial for these services, which must be easily operated on small screens and keyboards. Unfortunately, there are very few published studies on the evaluation of WAP devices and services on users. This paper presents a user study that evaluates two important interface design choices for WAP services (implementation of single-choice selection and navigation among the different cards of a WAP site) which have not been thoroughly investigated neither in the literature nor in design practice.

Keywords: evaluation, mobile phones, user interfaces, user study, WAP.

1 Introduction

WAP phones are a growing relevant part of the mobile market. For example, according to market analyses, at the midpoint of 2000 approximately 50% of the world's mobile phone shipments were WAP capable, and more than 40 million WAP browser-equipped handsets had been shipped [5]. Since then, the percentage of shipped WAP phones has steadily increased, and today most new models of cellular phones are WAP capable. This is spawning the need for and the development of services specifically tailored to mobile users employing WAP phones. As a result, the number of WAP sites keeps growing.

Usability is a crucial requirement for this kind of service, which must be easily operated by users on small screens and keyboards. Unfortunately, very little has been published on the usability of WAP phones and services, and the main focus of the few available papers (e.g., [6,10]) is not typically on the rigorous experimental evaluations demanded by HCI research. From a more general perspective, Abowd and Mynatt [1] point out how this generalizes to the whole body of work in Ubiquitous Computing, urging researchers to carry out thorough evaluations of systems for realistic human needs.

This situation has prompted us (and other researchers, e.g. [3,4,8]) to make user evaluation the main focus of our research on WAP interfaces. In particular, Buchanan et al. [3] compared three methods for displaying news headlines on a WAP phone emulator and more generally motivated the need for proper user-centered design approaches to make WAP services more effective and useful, while Ericsson et al. [4] present a usability evaluation of two commercial WAP browsers, employing real cellular phones. Nielsen [8] performed a field study using some real WAP services and pointing out their lack of usability, but the reliability of the study has been questioned (e.g., see [11]).

Our paper presents a user study that evaluates two important WAP user interface design choices (implementation of single-choice selection and navigation among cards) which have not been thoroughly investigated neither in the literature nor in design practice. Besides the specific results obtained, one of the contributions of the paper is also to present a case study of a carefully thought experimental procedure for evaluating different design choices for a WAP service.

The paper is organized as follows. In Section 2, we will present the alternative design choices considered in the study and further motivate the need for the experimental evaluation. Section 3 will thoroughly describe and discuss the different aspects of the user study we carried out, presenting the obtained results. Section 4 will conclude the paper with some final remarks on the study.

2 The Considered Design Alternatives

Current guidelines (e.g., [9]) on WAP service design tend to provide general principles that build on previous HCI knowledge (e.g., "always define a card for confirmation of potentially dangerous actions") or known technical limitations of WAP devices (e.g., "limit the length of labels to a maximum of 5 characters, otherwise some phones will truncate them"). These kind of guidelines can be formulated without the need of experimental backing, but leave many questions unanswered, e.g. about making design choices among the specific interaction styles supported by WML (language for content formatting in the WAP standard). Answering these questions requires thorough, complex, and time consuming experimental activity. In this paper, we consider two of these questions. The first concerns *navigation among cards* (NAC, hereinafter), i.e. the possible ways to move from one card to another in a WAP service. WML offers a choice between

two implementations of NAC: (i) *Links* (supported by the `<anchor>` tag): the user sees the possible destinations as underlined links in the current card, and clicks on them to move to the chosen card (see example in Fig. 1a), and (ii) *Action screen* (supported by the `<do>` tag): the user clicks on the "Options" key on the keyboard and sees the possible actions that can be taken in the current card listed on the full screen (see example in Fig. 2a).

The second question concerns *single-choice selection* (SCS, hereinafter), i.e. the display of a number of alternatives of which only one can be chosen. WML offers two possible implementations of SCS: (i) *List of links* (supported by the `<anchor>` tag): alternatives are displayed as underlined links and the user moves up/down to reach the desired one and click on it (see example in Fig. 1b), and (ii) *Selection screen* (supported by the `<select>` tag): the user sees her currently selected value between brackets, (s)he can accept it or click on it to see the alternatives listed on the screen and select another one (see example in Fig. 2b).

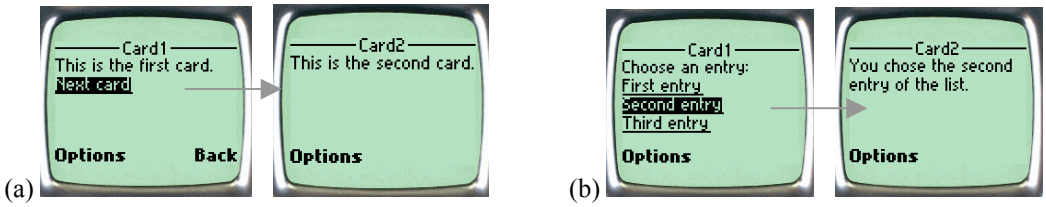


Fig. 1 (a) NAC implemented as a *Link*, (b) SCS implemented as a *List of links*.

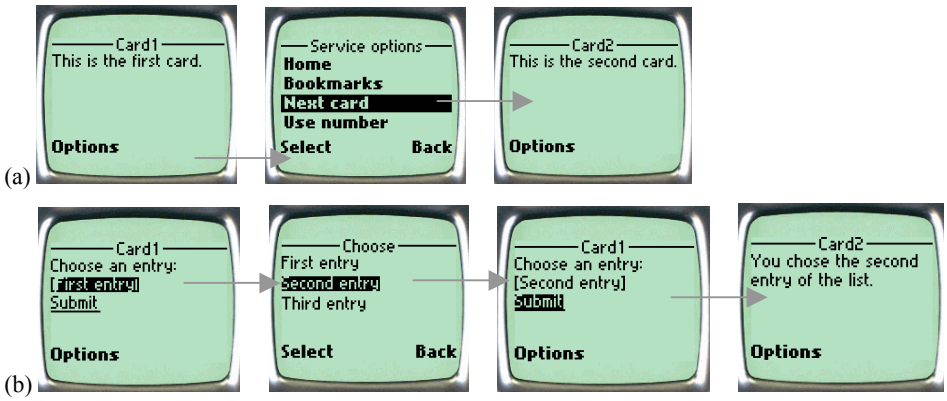
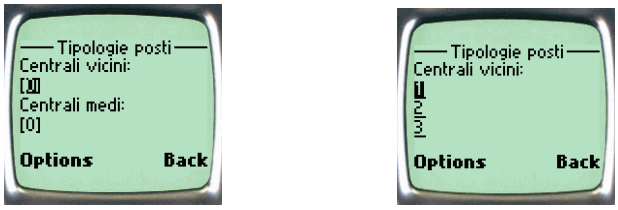


Fig. 2 (a) NAC implemented as an *Action screen*, (b) SCS implemented as a *Selection screen*.

Considering the trivial example displayed in Figures 1 and 2, one gets the impression that implementations in Fig. 1 (*Link* and *List of links*) are much simpler than those in Fig. 2 (*Action screen* and *Selection screen*). However, when more complex cards (typical of real WAP sites) are considered, one can get a very different impression. As an example, when a card contains more than one SCS

and there are many values to choose from in each SCS, a *Selection screen* implementation looks simpler than *List of links* because it allows the user to see the choices to be made and the selected values in a form-like screen that can be examined without scrolling too much, while *List of links* expands every selection and can require much more effort to determine the choices to be made and to reach the desired values. Figure 3 illustrates the situation with two alternative implementations of the same card in the CineWap service (described in Section 3.1) where a user is selecting the number and the position of some seats she wants to reserve in a cinema. With the *Selection screen* implementation (Figure 3a), the user sees the possible seat categories (central near the screen, central in the middle rows,…) and her selected values (how many seats she has reserved in each category) with less scrolling than *List of links* (Figure 3b).



(a) *Selection screen* (b) *List of links*

Fig. 3 Alternative implementations of the same card in the CineWap service.

Our focus on the two above described WAP interface elements is motivated by a number of reasons: (i) they are crucial in the development of any WAP site, and can deeply affect its usability, (ii) existing guidelines for WAP interface design (such as [9]) do not give criteria to choose among the described possibilities, (iii) frequently cited WAP references (such as [2]) stress that research is needed on providing guidelines for them, (iv) choosing the best implementation for these elements has been and is a topic of debate in mailing lists of WAP developers (such as [12,13]), witnessing the lack of design guidance and consensus about these aspects and the practical importance of studying the issue.

3 Experimental Evaluation

In the literature, the few user studies that concern WAP services are typically performed using WAP phone emulators (such as the Nokia or the Phone.com emulators) on a normal PC. We do not believe such kind of testing can provide reliable indications of user performance, for several reasons: (i) the user

experience is sensorially too different from the real use of the device (both from a visual and haptic point of view), (ii) the employed peripherals are too different both for the input and the output, (iii) the interaction styles are too different (pointer-mediated interaction with virtual controls vs. direct interaction with physical controls). To provide a practical example of unreliable data obtained through emulators, let us consider roller controls. To operate a roller on emulators, the user has to click with the mouse on a picture of the roller: clicks falling in the upper, central, or lower area of the picture will respectively roll up, press, or roll down the control. This may well make it more likely to inadvertently select the wrong action (e.g., moving the pointer only a few pixels can sometimes take it to the press area instead of the intended roll area and vice versa) with respect to a physical roller for which pressing, rolling up, and rolling down are different user motor actions. As a consequence, an interaction technique that heavily relies on the roller could artificially appear worse than an alternative technique based on keyboard buttons, if it is evaluated on a phone emulator.
For all the above described reasons, our experimental activity was carried out using a real WAP phone wirelessly connected to a full WAP service.
As explicitly indicated by research performed by mobile phones manufacturers [7], user efficiency and overall usability are critical design attributes in the mobile field. We thus devoted particular attention to their evaluation.

3.1 Subjects, Experimental Setting and Task

A total of 40 subjects (equally split between males and females) was involved in the experiment. The subject population was chosen to be representative of the general public, and to be diverse (in age, occupation, and ability with computers). Age ranged from 18 to 59, averaging at 29. Half of the subjects were students in different fields (Arts, Business, Foreign Languages, Medicine), while subjects in the other half held very different jobs (e.g, clerk, electrician, engineer, high school teacher, insurance agent, sales representative, secretary,…). The ability of subjects with computers was rated by asking them questions such as how frequently they use computers, since when, for what purposes: 3 subjects could be considered experts, 6 had no experience at all with computers, and the remaining 31 knew the basics of computer usage. With respect to Web experience, 15 subjects were not familiar with the Web, 12 were beginners, 7 intermediate, and 9 expert users of

Web sites. With respect to conventional mobile phones (not WAP-enabled) usage, 6 subjects had no experience at all, 18 used the basic features of a mobile phone, 16 used many advanced features, and the remaining 3 used every advanced feature. With respect to WAP phone experience, none of the subjects had used a WAP phone before.

The experiment was carried out using a Nokia 7110, connected wirelessly to a full WAP service. This allowed us to take the experimental activity out of our usual research lab, and move it into a common home environment that made subjects more at ease.

The considered WAP service (CineWap) is a reservation system that allows the user to get information about which movies are played in a city's cinemas and which seats are available at what times, making the desired reservations. The system offers also more advanced functionalities such as search (by title, genre, date, actor, director, cinema, rating), detailed descriptions of both movies and cinemas, ratings (users can give their evaluation of a movie they have seen), and alert service (users can be informed as soon as a specific movie opens in their city). Cards about specific movies, cinemas and shows are dynamically generated from a database: in particular, the database employed in the experiment contained 20 cinemas, 70 movies, and 4000 shows. We implemented the CineWap prototype following the WAP 1.2 standard, and using: (i) WML as a content formatting language, (ii) ASP for server-side scripting, (iii) Microsoft Access 2000 as a database management system. Client-side scripting has been completely avoided to take into account the limited CPU power of some current WAP-enabled phones. Currently available general usability guidelines for WAP services (e.g., usage of each WML tag, suggested constraints such as maximum length of cards, avoidance of browser-specific or phone-specific features,…) have been followed in designing the system. CineWap is particularly suited as a test service for the aspects considered in this paper, because it contains a large number of SCSs and a deep navigation structure.

The experimental task had to: (i) be representative of frequently occurring tasks in real WAP sites, and (ii) heavily involve the interface elements under evaluation. Therefore, we chose a search and selection task where the user had to find a pair of movies satisfying given requirements and reserve seats for a given show time.

Before carrying out the task, subjects were individually instructed (both orally and by showing interaction examples) about the functionalities of the WAP phone and the CineWap service. Fig. 4 shows a picture of the employed mobile phone, highlighting the purpose of its different controls in the considered task. Subjects were also allowed to spend some free time interacting with: (i) first, a very simple WAP site presenting the proper selection and navigation features, to become familiar with the controls and interaction styles, and then (ii) a training version of the CineWap site that employed the selection and navigation features which the subject had to later use in the experimental task.

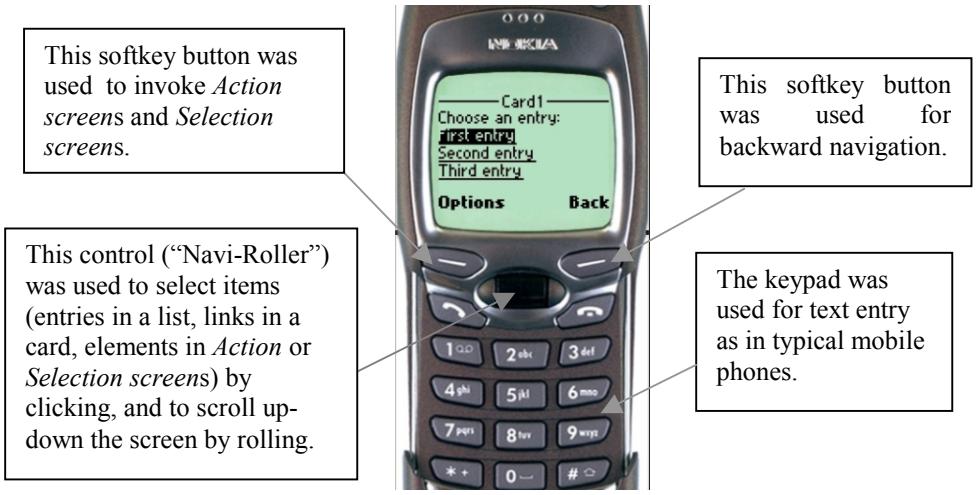


Fig. 4. Purpose of the different controls in the employed mobile phone.

3.2 Variables, Experiment Design, Task, and Hypotheses

Two independent variables are involved in the study: (i) implementation of NAC (with two levels: *Links* or *Action screen*), and (ii) implementation of SCS (with two levels: *List of links* or *Selection screen*). The following dependent variables were measured to characterize user efficiency and usability:

- *Percentage of task* completed in a given time, measured as the percentage of correctly visited cards (determined through software logging) among the set of cards necessary to complete the task. The given time interval was chosen with a pilot study on a few subjects that determined an average time to complete the task equal to 9 minutes.
- *Time needed* to complete the task (for subjects who were not able to complete the task, time needed coincides with the 9 minutes maximum allowed time). Since time spent can be influenced by variations in the connection speed, we used a video recording to precisely eliminate connection times (which

anyway did not show large variations). A digital video camera was set to record the entire phone and the fingers of the subject operating it. The camera was placed in such a way that the subject clearly felt that his/her face was not recorded. To make subjects further at ease, they were told that the purpose of video recording was to evaluate technical performance of the phone display (so, implicitly, that it was not meant to evaluate them).

- *Number of clicks* needed to complete the task.
- *Number of deviations* users took with respect to the shortest path of cards from the initial state to the state required by the task. This variable gives an indication of how much the interface facilitated the user in trying to reach the required state. A deviation was detected when the user visited a card (and possibly a number of following unnecessary cards) that was not prescribed by the shortest path of cards.
- *Total number of cards* visited in the WAP site to complete the task. This information integrates the previous one, specifying the length of the path of cards the user has actually followed to reach the required state, and was easily counted by software logging code.
- *Subjective evaluation* of the difficulty of navigating among cards and making selections, provided by users through a questionnaire.

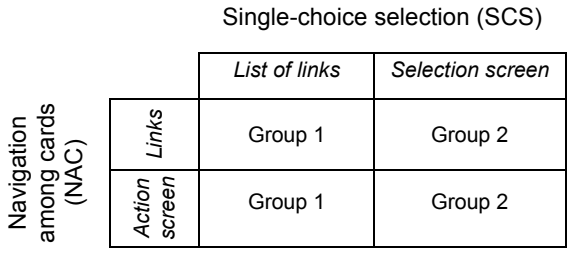


Fig. 5. Experiment design.

A 2 x 2 factorial design was adopted, considering the four possible combinations of the studied implementation choices. As shown by Fig. 5, subjects have been assigned to the four different conditions following a mixed design. First, according to a between-subjects design, subjects have been split into two groups of 20 subjects each, carefully assigning subjects to prevent significant differences in computer, Web, phone experience, and age between the two groups. Subjects in Group 1 employed only *List of links* for SCS, while subjects in Group 2 employed only *Selection screen* for SCS. Then, in each group, every subject performed the task in the two possible conditions for NAC according to a within-subjects design.

Therefore, two different versions of the task were needed and we designed them in such a way that they were of the same complexity and used the considered interface element in the same proportion. The *first version of the task* is organized into 3 subtasks:

1. Reserve seats for a specific movie. More precisely: (a) Search the movie whose title is "The sixth sense"; (b) Among the possible cinemas, choose the "Capitol" cinema; (c) Choose the 10PM show on the 9th of December 2000; (d) Reserve two side seats in the back rows.
2. Reserve seats for another specific movie. More precisely: (a) Search those movies whose director is "Zemeckis"; (b) Among the possible movies, choose the first one; (c) Among the possible cinemas, choose the one with the highest number of seats; (d) Choose the 8PM show on the 11th of December 2000; (e) Reserve one central seat in the front.
3. Cancel the reservation made for the first movie (see subtask 1).

The *second version of the task* has exactly the same structure; changes concern name of movie, cinema, show time, number and type of seats in subtask 1; search criteria (actor name), show time, number and type of seats in subtask 2.

For any experimental condition, the optimal path for both versions of the task required to visit the same number of cards in the WAP site (i.e., 31 cards) and encounter the same number of SCSs (i.e., a total of 9 SCSs). Figure 3 (described in Section 2) shows one of the visited cards for seat selection in its two possible implementations.

In each experimental condition, subjects were verbally introduced to the proper version of the task, which was also written on a sheet of paper that they kept during task execution. Subjects had a 5-minute break between the completion of the task in their first experimental condition and the start of the second one.

The assignment of subjects to version of the task and NAC conditions in the within-subjects part of the design was carried out following a counterbalancing scheme. More precisely, 25% of the subjects started with the *first* version of the task in the *Links* condition, 25% of the subjects started with the *first* version of the task in the *Action Screen* condition, 25% of the subjects started with the *second* version of the task in the *Links* condition, and 25% of the subjects started with the *second* version of the task in the *Action Screen* condition. In this way, we counterbalance learning effects and we also counterbalance the effects that might

arise from a possibly higher complexity of a version of the task over the other, in case the previously discussed effort to keep complexity constant might have left something unaccounted for.

Our hypotheses for the experiment were that: (i) *Links* provide better user performance than *Action screen* for NAC, and (ii) *List of links* provide better user performance than *Selection screen* for SCS. Hypothesizing these effects for the *number of clicks* variable is trivial, because navigating with *Action screens* and making choices with *Selection screen*s objectively requires more clicks (to invoke the corresponding screens). For the other dependent variables, we were instead motivated by the consideration that when an *Action screen* or *Selection screen* is displayed, the user temporarily loses sight and access to the contextual information provided in the card (s)he was examining. Therefore, we believe (s)he can become more easily disoriented (and require more mental effort for orientation) in navigating the WAP site with respect to *Links* and *List of links*.

3.7 Statistical Analysis and Experimental Results

A two-factor mixed-design analysis of variance (ANOVA) has been performed with implementation of NAC as the within-subjects variable, and implementation of SCS as the between-subjects variable. The results indicate that the effect is significant: for each dependent variable, a value of $p<0.01$ has been obtained. Means are given in Tables 1 and 2. All values in the two tables show that user performance was better in the *Links* condition for NAC, and in the *List of links* condition for SCS. The results in the percentage of completed task (both close to 100%) reflect the fact that the time interval we gave was sufficiently wide, so the majority of subjects was able to complete the task, and the other subjects were anyway close to completion. The large difference in the number of clicks is no surprise, since it is partly due to the nature of the different interaction techniques (as previously discussed). The findings we consider to be most interesting are those concerning the number of deviations and total number of visited cards. As shown by the two Tables, subjects in the *Action screen* and in the *Selection screen* conditions are more likely to deviate from the shortest path of cards that leads to task completion and to visit more cards than needed. This is consistent with our conjecture that having to invoke a separate screen (for navigation or selection purposes) while using a card is detrimental to user orientation.

In the questionnaire for subjective rating of difficulty, every subject was asked (with two questions) to rate how difficult it was navigating among cards in the *Links* condition and in the *Action Screen* condition, respectively. Then, every subject was also asked (with one question) how difficult it was to make selections. Ratings were on a 5-levels scale: very difficult, difficult, normal, easy, very easy. For analysis purposes, we mapped ratings from the qualitative scale of the questionnaires into integers. Table 3 shows the mapping and the obtained means. Results concerning NAC were analyzed using the Wilcoxon test (dependent samples), while results concerning SCS were analyzed using the Mann-Whitney test (independent samples). In both cases, a value of $p<0.01$ has been obtained. Subjective ratings confirm the results provided by objective measurements, clearly showing that users perceive a greater degree of difficulty in the *Action screen* and *Selection screen* conditions.

Table 1. Navigation among cards (NAC): means.

	Links	*Action screen*
Time needed	469 sec.	513 sec.
Percentage of task	97.4%	96.1%
Number of clicks	61.30	129.50
Number of deviations	0.75	1.85
Total number of cards	34.10	38.00

Table 2. Single-choice selection (SCS): means.

	List of links	*Selection screen*
Time needed	467 sec.	514 sec.
Percentage of task	97,7%	95,8%
Number of clicks	79.10	111.60
Number of deviations	0.77	1.90
Total number of cards	33.90	38.20

Table 3. Means of subjective ratings obtained with questionnaires.

Navigation Among Cards		**Single-choice Selection**	
Links	*Action screen*	*List of links*	*Selection screen*
4.07	2.70	3.65	2.75

(5=Very Easy, 4=Easy, 3=Normal, 2=Difficult, 1=Very Difficult)

4 Final Discussion

The user study presented in this paper has provided evidence that exploiting links for navigation and for single-choice selection purposes can significantly increase user efficiency and overall usability of services intended for WAP mobile phones, with respect to alternative solutions supported by WML.

In closing, we would like to add a word of caution about the generalization of the obtained results to other contexts. First, although our experiment was based on a representative example of current mobile phones, the market is likely to see the introduction of radically new designs with very different input/output modalities that could deeply affect user interaction. From this point of view, the study we presented should be repeated with representative examples of other classes of WAP-enabled

devices, if input/output is very different from the considered one. Some current market trends such as substituting the roller control with a thumb-operated micro-joystick should instead not drastically change the results: similarly to the navy-roller, the micro-joystick can be physically moved forward and backward to scroll and can be pushed to select; moreover, a possible increase/decrease in user performance due to the micro-joystick should affect every condition in the experiment similarly.

Second, we would like to stress the fact that we studied the behavior of *novice* users of WAP services, which could differ from that of *expert* users. The choice of focusing on novice users was motivated by the fact that WAP is a recently introduced technology for which attracting new users is a priority, and experts users are currently very few. However, studying expert users will become an important topic, especially when WAP sites will feel the need of providing the two categories of users with different interfaces.

References

1. Abowd G.D., Mynatt E.D.: Charting Past, Present, and Future Research in Ubiquitous Computing. ACM Transactions on Computer-Human Interaction **7** (March 2000) 29-58
2. Arehart C., Chidambaram N., Guruprasad S., et al.: Professional Wap, Wrox Press, Birmingham (2000)
3. Buchanan G., Jones M., Thimbleby H., Farrant S., Pazzani M. Improving mobile internet usability. Proc. 10th Internat. WWW Conf., ACM Press, New York (2001) 673-680
4. Ericsson T., Chincholle D., Goldstein M.: Both the Cellular Phone and the Service Impact WAP Usability. Joint Proc. of IHM 2001 and HCI 2001. Springer Verlag, Berlin (2001).
5. Gillott I.: Exploding the Myths of WAP. IDC Report, http://www.cio.com/analyst/112700_idc.html (2000)
6. Kaasinen E., Aaltonen M., Kolari J., Melakoski S., Laakko T.: Two Approaches to Bringing Internet Services to WAP Devices. Proc. 9th Internat. WWW Conf., Computer Networks Journal **33** (2000) 231-246
7. Jokela T., Pirkola J.: Using Quantitative Usability Goals: A Case Study about Development of a User Interface for a Cellular Phone. Proc. INTERACT 97. Chapman and Hall, London (1997)
8. Nielsen J.: WAP Field Study Findings. www.useit.com/alertbox/20001210.html (2000)
9. Openwave Systems.: GSM Application Style Guide. http://www.phone.com/pub/gsm900-1800.pdf (2000)
10. Schmidt A., Schroder H., Frick O.: WAP – Designing for small user interfaces. Proc. CHI2000 Conf. Human Factors in Computing Systems, Abstracts Volume. ACM Press, New York (2000) 187-188
11. WAP Forum.: Statement from the WAP Forum regarding a Nielsen Norman WAP Report. Press Release, http://www.wapforum.org/new/20010102244STA.htm (2001)
12. *wap-dev,* mailing list at wapwarp.com/wap-dev
13. *WML and WMLScript Programmers List,* mailing list at www.egroups.com