

Ion Juvina · Herre van Oostendorp

Individual differences and behavioral metrics involved in modeling web navigation

Published online: 7 January 2006
© Springer-Verlag 2006

Abstract This paper presents an empirical study aiming at investigating individual differences and behavioral metrics involved in modeling web navigation. Factors that have an influence on web navigation behavior were identified with the aid of task analysis, and their relevance in predicting task outcomes (performance, satisfaction, perceived disorientation) was tested with the aid of multiple regression analysis. Several types of navigation metrics were calculated based on web logging data and used as indicators of user characteristics and task outcomes. Results show that spatial-semantic cognitive mechanisms seem to be crucial in adequately performing web navigation tasks. The fact that user characteristics and task outcomes can be estimated with reasonable accuracy based on navigation metrics suggests the possibility of building adaptive navigation support in web applications.

Keywords Web navigation · Individual differences · Navigation metrics

1 Background

Navigation is a major part of user experience on the web [1]. This particular type of behavior is triggered by a specific type of applications, which has become very common nowadays, namely *web-based applications*. The user interface of these applications (web interface—WI—as called in [2]) has some characteristics that differentiates it from other types of interfaces, such as command language interface (CLI), graphic user interface (GUI), direct manipulation interface (DMI), interface

windows, icons, menus and pointing devices (WIMP), speech user interface (SUI), virtual reality (VR), etc. Unlike in GUI, DMI and WIMP interfaces, where mainly the functionality of an application is explored, WIs prompt the user to explore the domain knowledge. In fact, web users face two different interfaces:

- The browser interface, which remains consistent in daily use, and
- The site interface, which changes from site to site.

While the browser interface is rather easy to learn, it is impossible to provide adequate training on how to navigate through the many thousands of websites that the user may visit [1].

Web interfaces and the facility of navigation through large information spaces brought new problems for application designers and usability specialists; *cognitive overload* and *disorientation* are the main ones [3]. Interfaces have been traditionally designed with the function of providing users with information and means so that they can perform their tasks. In the case of WIs, this function has developed so much that it has almost become a burden for the user. Therefore, adequate ways to filter the information that is offered to the user and to guide navigation through the information space are necessary. The user must also be assisted in deciding what information is relevant, trustworthy, useful, etc. In order to achieve these functions, WIs must be aware of the user; in other words, they must incorporate a model of the user.

There is a vast amount of literature showing and analyzing individual differences involved in web navigation. Thus, Eveland and Dunwoody [4] notice that novices tend to make use of a linear structure in hypermedia systems when it is made available, while experts tend to navigate non-linearly. MacGregor [5] demonstrated that students who had greater *domain knowledge* evidenced more purposeful navigation and allocated time more variably to different information nodes when they were studying using hypertext

I. Juvina (✉) · H. van Oostendorp
Center for Content and Knowledge Engineering
Institute of Information and Computing Sciences,
Utrecht University, Padualaan 14, 3584 Utrecht, CH,
The Netherlands
E-mail: ion@cs.uu.nl
E-mail: here@cs.uu.nl

environments. Also, novices who possess less *domain knowledge* do not benefit from menu choices as much as experts [6]. *Spatial ability* is an important determinant of hypermedia navigation performance, as reported in several studies [7]. It has also been shown that individuals with low *spatial abilities* have difficulties in constructing, or do not use, a visual mental model of the space [8], and they are more directed to the semantic content [9]. Students with an internal *locus of control* are better able to structure their navigation and take advantage of hypertext learning environments [5]. Aging is associated with decreases in *working memory capacity* [10] and *computer confidence* [11]. Women report higher levels of spatial anxiety, which is negatively related to the orientation way-finding strategy [10].

Research aimed at modeling cognitive mechanisms involved in web-navigation is gaining increasing influence in the HCI community [12–16]. A cognitive model of Web navigation should be able to simulate the navigation behavior of real users. An example of such a model, called CoLiDeS, is proposed in [16]. It explains how users parse and comprehend the content of a Web page and then select what action to perform next. This model uses Latent semantic analysis (LSA) [17] to estimate the semantic similarity between user goals and semantic objects on web pages (e.g., link anchors). CoLiDeS constitutes the theoretical base of a usability evaluation method, called Cognitive Walkthrough for the Web (CWW) [18], which is used to identify and repair usability problems related to navigation in web sites.

A related line of research aims at modeling the user's navigation behavior in order to provide adaptive navigation support in web applications [19]. A user model can include (relatively) stable characteristics such as gender, age, education level, and dynamic (changing) characteristics such as goals and preferences. Stable characteristics do not pose any difficult problem for the designer of a personalized web application. The dynamic navigator's model is more challenging and more useful for the goals of personalization. A dynamic navigator's model could include:

- Syntactic information about navigation behavior (which links are followed, in which order, how does the navigation graph look, e.g., linear or nonlinear).
- Semantic information (what is the meaning of the information that the user encountered during navigation, which of this information was processed/found relevant by the user).
- Pragmatic information (what is the purpose of the user in using that information, what are the user's goals and tasks).

The distinction between syntactic/semantic/pragmatic information applied in the analysis of web navigation behavior is analogous to the same distinction in

the field of linguistics. In the context of modeling web navigation as discussed in this paper, 'syntactic' means structural, topologic information, 'semantic' refers to the content of visited pages, and 'pragmatic' information indicates what are the reasons and gains of visiting certain pages.

1.1 Research objectives and questions

The study presented in this paper aimed at building a model of web navigation with applicability in web usability and design of personalized web applications.

The initial step was to identify factors that were able to predict task outcomes such as performance, satisfaction and reliability. Some of these factors were person-related (user characteristics) and others were interface- and context-related.

Subsequently, navigation data was used to estimate person-related factors (user characteristics) and predict task outcomes. The focus was on data automatically recorded as a byproduct of a navigation session (web logging data), because it was easy to collect such data in real time and unobtrusively.

A review of the issues that were considered in this study is presented in Fig. 1. Research questions about these issues were as follows:

- Are the hypothesized factors indeed significant as predictors of task outcomes? Which ones are the best?
- What is the relative importance of each factor in predicting task outcomes?
- How well can each of the task outcomes be predicted?
- Is it possible to predict user characteristics based on navigation behavior? For example, how accurately *spatial ability* can be predicted based on navigation metrics?
- Is it possible to predict task outcomes based on navigation behavior? For example, how accurately can user's *perceived disorientation* be predicted based on navigation metrics?
- How accurately can task outcomes be predicted based on both user characteristics and navigation metrics?

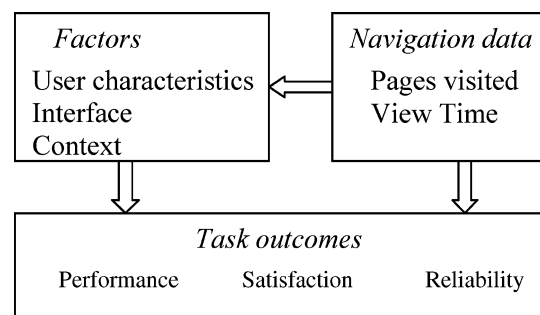


Fig. 1 Overview of issues considered in this study

2 Methodology

2.1 Task analysis

Published research shows that *domain knowledge* (knowledge that is needed to successfully perform tasks in a particular domain) is a key factor in successful web applications [7], and other factors could also be domain dependent. Consequently, a task domain had to be defined in order to capture the influence of those factors that are domain specific.

A valuable source of insight into what types of activities significantly impact people's decisions and actions was found to be the survey conducted by Morrison et al. [22]. Looking at the collection of incidents mentioned in [22], one can notice that there are some activities that seem to be frequent and specific enough to constitute a domain, and that this type of activities has received insufficient attention in research and application so far. The following incident is an example:

I accessed Netscape's financial site to check my credit card balance and how long it would take to pay it off. I'm now MUCH more fiscally aware of my spending habits and am trying to pay off my balance more actively.

We call this domain Web-assisted Personal Finance (WAPF). It includes using the web to setup personal financial goals, keep a personal budget, decide to save or invest, do financial transactions, finance life events such as studying, etc. Three websites were used in this study. Two of them are dedicated to WAPF, and provide users with advice and tools (such as planners, calculators, educators) to deal with their financial problems. The third one, an e-commerce website, was used as a reference, being known as a reasonably well-designed web application.

An exploratory task analysis within this domain allowed to understand which were the most relevant success factors and to build a hypothetical model [48]. Some results of task analysis that determined the composition of the hypothetical model are:

- Some subjects were capable of deploying a fast, elaborated and effective web navigation behavior. Consequently, a factor concerning Internet expertise was considered.
- Although tasks were conceived in such a way as to require as little previous knowledge as possible, it was noticed that a certain familiarity of users with the financial domain was an advantage.
- Spatial ability was included in the hypothetical model based on the high frequency of spatial terms used in subjects' verbalizations, even when they were dealing with completely non-spatial issues. Examples of verbalizations with spatial connotation include: "where am I", "let's go in another place", "I'm stuck in these analyzers", "I saw it somewhere".

Another goal of task analysis was to collect realistic task instances. Real life tasks do not require only finding information, but also problem solving and decision-making [23]. Tasks in WAPF include:

- Information search (e.g., "What is the definition of financial goal?");
- Personal life planning (e.g., "Setup a personal budget");
- Problem solving (e.g., "How much do I need to save monthly in order to buy a car in 4 years?");
- Personal decision-making (e.g., "What kind of car can I afford?").

2.2 Variables and indicators

2.2.1 Task outcomes

Criteria for task outcomes were specified during task analysis. The intention was to find a small number of criteria to cover as many task outcomes as possible. *Effectiveness*, *efficiency* and *satisfaction* were taken from ISO9241-11 [24] and effectiveness and efficiency were grouped under the label *performance*. *Performance* denotes task success (effectiveness) obtained with minimum resources (efficiency). Satisfaction refers to users' affective experience toward task execution and task results.

Besides *performance* and *satisfaction*, another criterion was considered necessary to cover the undesirable aspects of task outcomes. There is a vast literature showing that models of human performance are incomplete if they consider only correct performance and neglect human error or, more general, human fallibility. For example, Reason [25] states that correct performance and error are like *active* and *passive* sides of a cognitive balance; each *debit* has a corresponding *credit*. For instance, skills development increases performance but also the risk of error, by turning off the conscious control mechanisms. In the field of human-computer interaction, van Oostendorp and Walbeehm [26] argue for the necessity of (and propose some modeling techniques for) considering errors, inefficiency and problem-solving processes in modeling human behavior in interaction with direct manipulation interfaces. Since the work presented in this paper takes into account not only errors, but also other undesirable aspects of task execution, such as stress, cognitive workload, disorientation, frustration, violations of privacy, etc., a more generic term was chosen, namely *reliability*, which is the antonym of 'fallibility'. In this context, reliability refers to avoiding or minimizing negative outcomes of task execution (see Sect. 2.3. for an operationalization of reliability).

2.2.2 Predictors

As already discussed in Sect. 1, potential predictors include user characteristics, interface and context factors,

and navigation metrics. They will be described in more detail in this section.

2.2.2.1 User characteristics Some user characteristics were hypothesized (based on task analysis and previous research) to have an influence on task outcomes. They were grouped, on a conceptual basis, into cognitive, affective, conative and demographic factors.

The distinction between cognition, affection and conation is well-founded in psychology; all human behavior involves some mixture of all three aspects [29–32]. *Cognition* refers to the process of coming to know and understand, i.e., the process of encoding, storing, processing, and retrieving information. It is generally associated with the question of “what” (e.g., what happened, what is going on now, what is the meaning of that information). *Affect* refers to the emotional interpretation of perceptions, information, or knowledge. It is generally associated with one’s attachment (positive or negative) to people, objects, ideas, etc., and relates to the question “How do I feel about this knowledge or information?”. *Conation* etymologically comes from the Latin verb *conare*, meaning to strive. It refers to the connection of knowledge and affect to behaviour, and is associated with the issue of “why”. It is the personal, intentional, deliberate, goal-oriented, or striving component of motivation, the proactive (as opposed to reactive or habitual) aspect of behavior [33]. It is closely associated with the concept of volition, defined as the use of will, or the freedom to make choices about what to do [34]. Some of the conative issues one faces daily are:

- What are my intentions and goals?
- What am I going to do?
- What are my plans and commitments?

Conation is absolutely critical when an individual is successfully engaged in self-direction and self-regulation [35]. Age and gender of participants were considered as demographic factors.

2.2.2.2 Interface and context factors Given the scope of this study, only a few of the interface and context factors that could have an influence on task outcomes could be investigated. Others were randomized or kept constant. By choosing three different websites to be used as research material, factors pertaining to site structure or interface design were randomized. Sites’ usability was explicitly measured as an interface factor.

With regard to different usage contexts that could have an influence on task outcomes, we kept as constant as possible the room, the type of computer and all the other contextual factors that could influence users’ navigation behavior, except the factor Time constraints that was experimentally manipulated. By trying to induce the feeling of time pressure in a half of the involved

subjects, the conducted study attempted to simulate part of the ‘mobile context’ of web navigation [27].

2.2.2.3 Navigation metrics An important amount of behavioral data about web navigation is available in web logs. Besides its availability, using this data in modeling web navigation is justified when one has the aim of personalization in mind: web logging data is collected unobtrusively and in real time, analysis of this data can be automated, and adaptive reactions of the application can be based on analysis results.

Interaction events that can be logged during a navigation session are quite numerous: page downloads, view time, use of buttons, etc. Some data about the web structure being navigated is also available: page title/URL, number of words per page, number of outgoing/incoming links etc. This constitutes the raw data of web navigation, which progresses toward information and knowledge via analytic and interpretational undertakings during the modeling process.

A number of analyses can be performed on the raw data in order to extract some useful *information* out of this data. The results of these analyses are referred to as *navigation metrics*. Extracting information out of navigation data by the aid of various navigation metrics is the way toward acquiring *knowledge* about the user.

Different types of information can be derived from navigation data: syntactic, semantic and pragmatic (see Sect. 1 for a description of these types). Within this study, based on raw web-logging data, two types of syntactic metrics (first-order and second-order metrics) and one semantic metric (path adequacy) were calculated as presented below. These metrics are used as predictors in the analyses presented in Sect. 3.

The raw data consisted of:

- Interaction events. For each *navigation session*, the following data was collected: page visited and link followed; time of visit; page load time and length of visit; navigation action (e.g., link, address bar, refresh, back button).
- Site structure data. For each *page* the following data was collected: URL and host; size in bytes; number of words and images; number of outgoing links; title and author; source code. For each *link* the following data was collected: source and target page; text associated with the link; type of link (internal, external).

First-order metrics A first set of navigation metrics was labeled first-order metrics, because it was derived directly from the raw data, without taking into consideration any usage matter. For example, average connected distance (ACD) was calculated independent of back button use (BBU), and did not take into consideration the fact that low values on ACD were associated with high values on BBU and vice-versa ($r = -0.49$). This latter information was used in calculating second-order metrics.

After successive trials, a number of 19 metrics were selected to be used in further analyses. They are briefly described below (Table 1). For a more detailed discussion about these metrics see [27].

Second-order metrics As it is most likely that patterns in the first-order metrics occur quite often simultaneously, second-order navigation metrics—linear combinations of the first order metrics—were calculated. They are described in Sect. 3.2.

Path adequacy A semantic metric, called Path adequacy, was calculated based on navigation data and the task descriptions that subjects were provided with at the beginning of the navigation session. This metric is also described in Sect. 3.2.

2.3 Operationalization

The operationalization of the hypothetical model [48] consisted in expressing each factor in measurable variables and indicators.

Thus, the cognitive factor *expertise* was first divided in *Internet expertise* and *Finance expertise*. An *Internet expertise* measure was constructed based on users self reported frequency of Internet use and their self-assessed level of knowledge and skills in web navigation. *Finance expertise* was measured with items such as: “Have you

ever used a personal finance website (Yahoo Finance, MSN Money etc.)?”.

The variables, *spatial ability*, episodic memory, and *working memory* were measured with computerized cognitive tests provided by TNO—Human Factors Institute. The ‘Spatial ability test’ used the classical mental rotation task, and the spatial ability score was the number of correct solutions obtained by rotating three-dimensional objects (correct matches between objects and their rotated equivalents). The ‘Episodic memory test’ presented three lists of 60 images each; the participants had to loudly name the images in the first two lists; between lists 2 and 3 there was a distraction task (we used the ‘spatial ability test’ as a distraction task, to efficiently use the testing time); list 3 contained images that were presented before in lists 1 and 2 together with new images; the participants had to recognize the images that were presented in list 1. The ‘Working memory test’ used a reading span task [36]: subjects were presented with series of phrases, the size of series increasing progressively from two to seven phrases; the participants were asked to loudly read the phrases and try to understand their content; after each series, the participants were asked to recall the last word of each phrase in that particular series; for one random phrase in the series participants were asked to fill in two missing words, to ensure that they really treated the whole content and not only the last words. The working

Table 1 First-order metrics

Metric	Short label	Description
Path length	<i>Pathleng</i>	The number of pages visited during the navigation session
Relative amount of revisits	<i>Revisits</i>	The probability that any URL visited is a repeat of a previous visit
Return rate	<i>Return</i>	The average number of times that a page was revisited
Back button usage	<i>Backbutton</i>	The percentage of back button clicks among the navigation actions
Relative amount of visits to the homepage	<i>Homepage</i>	Amount of visits to the homepage; “relative” refers to a correction based on path length
Average view time	<i>Meanview</i>	The total time spent on viewing pages divided by path length
Median view time	<i>Medview</i>	The median of the view times spent on every page
Difference between the average and the median view time	<i>Difview</i>	The extent at which the average view time is overly influenced by a minority of pages that are viewed carefully
Deviation in view time	<i>Devview</i>	It indicates how much the view time varies between pages
Time spent on larger pages	<i>Viewlarg</i>	Time spent on pages with number of words larger than the average plus the standard deviation
Time spent on smaller pages	<i>Viewsmall</i>	Time spent on pages with number of words smaller than the average minus the standard deviation
Time spent on index pages	<i>Viewindx</i>	Time spent on pages with a large number of outgoing links
Time spent on reference pages	<i>Viewref</i>	Time spent on pages with a small number of outgoing links
Relative Structural Complexity (‘fan degree’)	<i>Fandeg</i>	The ratio between the number of links followed and the number of distinct pages visited. A large fan degree indicates that users tend to follow more than one link from each page
Number of cycles	<i>Cycles</i>	The difference between the number of pages and the number of links. It indicates the number of non-linear navigation steps
Path density	<i>Density</i>	It measures at what extent users make use of all possible links of the site structure. A high path density indicates that a user makes use of short navigation sequences and returns to visited pages
Compactness	<i>Compact</i>	It is a measure similar to path density except that it takes into account the actual site structure (not all possible links in the site)
Stratum	<i>Stratum</i>	It is a metric designed to capture the linear ordering of user navigation. A lower stratum means a lower distance between any two nodes and indicates less linear navigation [28]
Average connected distance	<i>ACD</i>	It indicates the average length of a path between any two connected pages

memory score was calculated based on correctness of recalls. This test is more complete and more adequate than digit span tests for working memory capacity, since it takes into consideration not only information storage but also information processing that is normally associated with working memory capacity.

Locus of control refers to the individual's belief regarding the causes of his or her experiences, and those factors to which an individual attributes his or her successes and failures. Research shows that users with an internal locus of control are better able to structure their navigation and take advantage of hyperspace features [5]. Locus of control was measured with a 20-item scale [37]. The *sequential-holistic* cognitive style was measured with items such as: "I like to break down large problems into smaller steps" and "I like to look at the big picture" [38].

A measure of users' affective disposition at the beginning of the navigation session was built based on users ratings of different affective states that they considered appropriate to describe their current disposition. Subsequently, users' ratings were factor analyzed and grouped in three basic *moods*. Thus, *active mood* was composed by the following affective states: determined, calm, alert/vigilant, sluggish/lethargic/lazy (negative sign), and blue/depressed (negative sign); *Enthusiastic mood* was mainly composed of the enthusiastic, excited, and strong states; and *irritable mood* contained mainly the states irritable, sluggish (lethargic, lazy), nervous, sleepy, and relaxed (negative sign).

Participants' *propensity toward trust* [39] was measured with items such as: "People always can be trusted" and "People always take care only of themselves".

The factor called 'Motivation' was included in the model based on observations during the experiment and inspections of students' answers to questionnaires items. A dichotomous variable that differentiates between participants from Utrecht University and Twente University was initially recorded just to check for sampling errors. Afterwards, it was noticed that the students in the two universities reported consistently different types of interests, e.g., students from Utrecht University declared higher levels of interests in entertainment and personal development, whereas students from Twente University declared higher levels of interests in personal and professional businesses. This variable was hypothesized to pertain to students' motivation and goal orientation. The differences between the two groups of students (Utrecht vs. Twente) seemed similar to the difference between mastery and performance goal orientation. *Mastery* oriented students perceive new tasks as an opportunity to learn or to acquire new skills, whereas *performance* oriented students perceive tasks as opportunities to demonstrate already existing competence and skills [40]. This hypothesis must be checked in further research, but, for this study, the new dichotomous variable was used with the temporary label "Motivation".

Self-efficacy was measured with a questionnaire, adapted from [41], containing items such as "I could perform better using these websites if I had a lot of time to complete the job for which the sites were provided".

The *interests* factor of the hypothetical model was operationalized based on principal component analysis. Participants were asked to indicate for what purposes they use the Internet. Answers were factor analyzed, and the two components that resulted were called 'Interest entertainment and personal development' (for brevity, *interest entertainment*) and 'Interest personal and professional business' (for brevity, *interest business*), respectively.

Perceived usability of the three sites used in the study was measured with a selection of items from questionnaires [42] and [43], consisting of items such as: "It was easy to use this website" and "I could effectively complete my tasks using this website".

The factor *time constraints* was experimentally manipulated. Half of the subjects (15) were instructed that only 30 min was available to complete the navigation tasks, while the other half did not receive any time indication. In fact, all subjects were given a maximum of 40 min to execute the navigation tasks. No clock or other time indication was available.

The criterion *performance* was operationalized only in 'effectiveness' (attaining of task goal). The degree of success for each task was rated from 0 to 4 based on correctness and completeness of answers. The 'Efficiency' side of performance was not directly considered because the time of the navigation sessions was kept constant for all subjects. In this case, 'efficiency' of task execution is implicitly considered in 'effectiveness'. Other non-temporal metrics of efficiency (e.g., number of steps taken in solving a task) were more or less captured by some of the navigation metrics (e.g., path length); including them as criteria would have artificially inflated the predictive power of the model.

Satisfaction as a criterion was measured by items such as "It was an interesting experience to perform these tasks", and "Overall, working to accomplish these tasks was satisfying". This criterion did not refer to the satisfaction of users toward the websites used; the latter was captured by the *usability* factor. By separating satisfaction toward the tools used (web sites) from satisfaction toward task execution and results we aimed to avoid the 'common measure bias' described by Takeishi [49].

Reliability as a criterion was operationalized in this study by variables *perceived disorientation* and *frustration*. *Perceived disorientation* was measured with items adapted from [44], such as "It was difficult to find the information I needed on this site", and "It was difficult to find my position after navigating for a while". *Frustration* was measured with items such as "I felt frustrated when I encountered difficulties in completing the tasks", and "I felt angry when I couldn't find what I needed to complete the tasks".

2.4 Subjects and procedure

The study was run with 30 participants in a single session, lasting approximately 2.5 h. 15 participants (7 females and 8 males) were registered as students in the Information Management Department of Twente University, and the other 15 participants (8 females and 7 males) were students in the Information Science Department of Utrecht University. Participants were selected randomly out of students' catalogues of both universities. Half of the participants were randomly assigned to the 'Time constraints' condition in which the participants were instructed to finish the navigation tasks in 30 min.

The first part of the sessions was dedicated to questionnaires and cognitive tests aimed at measuring user characteristics. The second part consisted in execution of web navigation tasks. This part lasted maximum 40 min for all participants (including those in the 'time constraints' condition). No clock was available, participants were asked to put away their wristwatches, and the computer clock was disabled. During the navigation task, navigation behavior and task performance were recorded. Subjects were informed that their navigation behavior was recorded. Task performance was recorded by the participants on a dedicated form and coded afterwards by the experimenter. The third part of the sessions consisted of administration of usability and satisfaction questionnaires. Each participant received a compensation of Euro 20 at the end of the session.

3 Results

Results are presented in the same order as suggested in the overview of the issues considered in this study (Fig. 1). Multiple linear regression analysis was used to investigate the significance of hypothetic factors in predicting task outcomes, as well as the possibility of using navigation metrics as estimates of user characteristics and predictors of task outcomes. Including predictors in regression models was based on the *stepwise method*, thus the predictive power must be seen as the best one can get with the minimum number of predictors. The input of regression analysis is composed of predictors and criteria described in Sect. 2.2. After each analysis, the outcomes of the stepwise procedure were summarized in tables presenting:

- The criterion (Dependent variable);
- The proportion of variance in criterion explained by the significant predictors (R^2);
- The predictors retained at the end of the stepwise method as significant;
- The relative importance (beta coefficient) of each significant predictor.

Predictors found to be not significant (excluded from the model) are not listed with every analysis but they can be easily recovered from the list of predictors presented in Sect. 2.2.2.

3.1 Predicting task outcomes based on hypothesized factors

All task outcomes could be predicted based on a limited number of predictors with various effect sizes. According to Cohen [45], the effect size for regression is calculated with the following formula $ES^2 = R^2 / (1 - R^2)$. An effect size of 0.02 is considered a small effect, 0.15 a medium one, and 0.35 a large one. In our case (Table 2), the smallest multiple R^2 (0.22) corresponds to a medium-large effect size.

Task performance was best predicted by *spatial ability* and *domain (finance) expertise*. In other words, the user ability to represent the structure of the sites and their domain knowledge were the most important determinants of task success. *Satisfaction* was best predicted by *motivation*, *usability* and *interest in business*. Users who were motivated, perceiving the websites as usable and not interested in business were more likely to be satisfied with task completion. *Business interest* negatively correlated with *satisfaction* ($r = -0.38$). A possible interpretation is that subjects with interests in personal and professional business have higher expectations and they are more vulnerable to be dissatisfied when task execution and results do not meet their expectations. *Disorientation* was best predicted by *usability* and *working memory*. Low working memory capacity and low perceived usability were associated with increased probability of users perceived disorientation. *Frustration* was predicted by *time constraints*. Users in the 'time constraints' condition reported a higher level of frustration than users in the control condition.

3.2 Predicting user characteristics based on navigation metrics

As presented in Sect. 2.2.2.3, several types of navigation metrics were calculated: *first-order*, second-order (navigation styles and reading time) and a semantic metric called *path adequacy*. Few details about how second order and semantic metrics were calculated are presented below.

Two different data analysis approaches were employed in deriving second-order navigation metrics: *unsupervised learning* (principal component analysis) and *supervised learning* (regression). The term "learning" has only a statistical connotation here. It means deriving new information based on existing information. In the unsupervised learning approach, only patterns of covariance in the first-order metrics were considered,

Table 2 Predictions of task outcomes based on the hypothetic factors

Task outcome	R^2	Predictors	Beta
Performance	0.39	Spatial ability	0.496
		Finance expertise	0.385
Satisfaction	0.67	Motivation	0.612
		Usability	0.506
Perceived disorientation	0.42	Interest business	-0.319
		Usability	-0.505
Frustration	0.22	Working Memory	-0.344
		Time constraints	0.471

regardless of any outside criteria (i.e., other independent measures, e.g., *task performance*). The second-order metrics resulted in this way were called *navigation styles*. They were completely specified (numerically) by first-order metrics. However, interpreting their meaning and labeling them was based on their correlations with user characteristics and task outcomes. In the supervised learning approach, the task outcomes defined in Sects 2.2.1 and 2.3 were used as outside criteria in the attempt to combine the first-order metrics. A second-order metric was derived in this way and was called *reading time*. It is a combination of several view time metrics weighted in such a way as to ensure a significant correlation with *task performance*.

3.2.1 Navigation Styles

A principal component analysis with *equamax* rotation was run on the 19 first-order metrics presented above. A 4-components solution that accounted for 85.95% of the initial variance has been selected. Each component accounted for 27.3, 23.8, 22.8, and 12.0 % of variance, respectively. Component loadings in first-order metrics and the correlations between factors and user characteristics and task outcomes were used to interpret the content of each factor in terms of navigation styles, as follows.

Component 1. Flimsy navigation High scores on this component were associated with small number of pages visited, high density, high view time per page, low average connected distance, low number of cycles, high rate of home page visiting, and high frequency of back button use. This appears to be a parsimonious navigation style. The navigation path was not very elaborated, most of the navigation taking place around the homepage. Time was spent in processing content instead of figuring out the hyperstructure that showed where the relevant information was. A high score on the flimsy navigation style was associated with low Internet expertise ($r = -0.5$), low active mood ($r = 0.48$), low working memory ($r = -0.38$), external locus of control ($r = -0.37$), and high perceived disorientation ($r = 0.46$).

Component 2. Content focus This component grouped together all the view-time metrics, which basically indicated that there was a general consistency in users' view-time allocation. In other words, users were con-

sistently 'slow' or 'fast'. High values on this component indicate high view time on a rather small set of pages. Within this style, navigation was a means for reading. Users' goal was to find those pages that ought to be read and then read them. This style was not associated with any user characteristics or task outcomes.

Component 3. Laborious navigation High scores on this component were associated with high number of links followed per page, high revisitation rate, high number of cycles, high returning rate, high use of back button, high density, high number of pages visited, low average connected distance (short returns). This style involved intensive use of navigational infrastructure provided by the site. Users seemed to employ a trial and error strategy. They followed links just to see if they were useful or not. They figured out quite fast when paths were not leading towards their goal and came back. Revisits were quite numerous but they were not redundant: once a page was revisited, a different link was followed, it was just another trial. This navigation style was associated with high episodic memory ($r = 0.49$), low spatial ability ($r = -0.40$), and low interest in entertainment ($r = -0.38$). This style indicates the type of revisitation that does not relate to disorientation. The user needed to look around for a while until s/he had a good representation of the site structure, because s/he had a weak spatial ability. Her/his memory though prevented her/him of making redundant revisits. This component shows how people compensate for the lack of spatial ability by effort and memory, and do not necessarily decrease performance (no correlation with task performance was found, although spatial ability and task performance were positively correlated). It also shows why revisitation is not always associated with disorientation.

Component 4. Divergent navigation High scores on this component were associated with low compactness, high stratum, low homepage use, high connected distance (long returns). This navigation style was rather explorative. Users were not that eager to revisit pages but rather to explore new directions. This navigation style was only associated with a high propensity to trust ($r = 0.43$).

Reading Time Another second order-metric was constructed by trying to get meaningful combinations of first-order metrics that would significantly correlate with task outcomes. One such trial that proved to be successful is the following:

$$\begin{aligned} \text{Readtime} = & 230.1 + 4.2 * \text{viewlarg} + 1.56 * \text{viewref} \\ & - 3.73 * \text{viewsmal} - 4.77 * \text{viewindx} \\ & + 2.04 * \text{devview}. \end{aligned} \quad (1)$$

It was obtained by running a stepwise regression analysis with *task performance* as dependent variable and all the first-order metrics as independent variables. The significant predictors selected by the stepwise procedure

(*viewlarg*, *viewref*, *viewsmall*, *viewindx*, and *devview*) and their corresponding *b*-weights together with the intercept (230.1) were used to build the Equation (1) (see above). A unique score was calculated as a linear combination of the significant first-order metrics. It was labeled *reading time* since it positively weighed view time on large and reference pages presumed to be content pages ($b=4.2$ and $b=1.56$, respectively) and negatively weighed view time on small and index pages ($b=-3.73$ and $b=-4.77$, respectively). High *reading time* was significantly associated with low *flimsy navigation* ($r=-0.5$), low *divergent navigation* ($r=-0.36$), high *Internet expertise* ($r=0.35$), high *performance* ($r=0.43$), and low *disorientation* ($r=-0.39$). *Reading time* can be interpreted as an efficient navigation style in which view time is higher on large and reference pages (since they require reading) and lower on small and index pages (since they only require scanning).

Path adequacy This semantic metric was calculated as a measure of semantic similarity between a navigation path and a task description. A navigation path was considered to be a concatenation of semantic objects that the user has encountered in her/his way toward a specific location. As semantic objects, one can consider link anchors, page titles, page contents, URLs, clickable icons, banners and images etc. Navigation paths composed of page titles have been used in the study.

For example, if the user visited the pages titled *Should I finance or pay cash for a vehicle? Calculators*, *How much will my vehicle payments be? Calculators*, *Glossary*, and *What vehicle can I afford? Calculators*, then his/her navigation path was represented as a string of all words in these titles: *<should, I, finance, or, pay, cash, for, a, vehicle, calculators, how, much, will, my, vehicle, payments, be, calculators, glossary, what, vehicle, can, I, afford, calculators>*.

Navigation paths were compared with task descriptions. The following is an example of task description:

Suppose you want to buy a car in 2 years. You have already saved \$ 500. How much do you need to save on a monthly basis in order to make a down payment of \$ 8000 for the car? Assume that the savings and tax rates are as listed. What is the most expensive car you can afford if you will be able to pay 40 monthly payments of at most \$150 after the down payment?

In order to measure the semantic similarity between navigation paths and task descriptions the technique called Latent semantic analysis [17] was used. High *path adequacy* was significantly associated with low *return rate* ($r=-0.48$), high *spatial ability* ($r=0.36$), high *self-efficacy* ($r=0.40$), and high *performance* ($r=0.47$).

Table 3 shows that a considerable number of user characteristics could be predicted with reasonable accuracy based on navigation metrics. For example,

25% of the variance in *Internet expertise* was predicted based on the *flimsy navigation style*. This result has a direct implication for providing adaptive support in web navigation. As *Internet expertise* is virtually impossible to measure in real-time use of a web application, it can be estimated based on user's navigation path.

3.3 Predicting task outcomes based on navigation metrics

Two of the task outcomes, *performance* and *disorientation*, were significantly predicted based on second-order and semantic metrics (Table 4). These results showed that second-order and semantic navigation metrics were better predictors of task outcomes than first-order metrics. There were no significant predictions for *satisfaction* and *frustration*.

3.4 Predicting task outcomes based on hypothetical factors and navigation metrics

When both hypothetical factors and navigation metrics (see Fig. 1) were entered in the regression analysis, the selection of the most significant predictors generated slightly different results but these results were consistent with the previous ones (Table 5). For example, *number of cycles* appeared among the significant predictors of *disorientation* instead of *flimsy navigation*. But *cycles* was highly correlated with *flimsy navigation*, actually it was one of its most important components. Two first-order metrics appeared as significant predictors of *satisfaction*, which can be interpreted as follows: the amount of variance in *satisfaction* left unexplained by *motivation*, usability and *interest for business* could be explained by *view time per large pages* and *view time per small pages*. Note that the beta coefficients had opposite signs: users were more satisfied when they spent relatively long time on large pages and relatively short time on small pages than vice versa. In other words, spending relatively long time on large pages (presumed to be content pages) determined that part of *satisfaction* in task completion that was not determined by *motivation*, interest for business and *sites usability*.

Another effect of considering factors and metrics together as predictors of task outcomes was a considerable increase in predictive power: R^2 was higher than 0.50 for all criteria except *frustration* (the average increase of R^2 is 0.133). This result showed that user characteristics, interface and context factors, on one hand, and navigation metrics, on the other hand, were rather complementary in predicting task outcomes. In other words, navigation metrics could not completely explain task outcomes by themselves, but they were able to indicate facets of task outcomes that were not explained by the hypothesized factors.

Table 3 Predictions of user characteristics based on navigation metrics

Characteristic	R^2	Predictors	Beta
Internet expertise	0.25	Flimsy navigation	-0.50
Spatial ability	0.195	Revisits	-0.442
Episodic memory	0.245	Laborious navigation	0.495
Working memory	0.28	Flimsy navigation	-0.83
		Median view time	0.58
Locus of control	0.398	Flimsy navigation	-0.704
		View reference	0.874
		Deviation view	-0.553
Active mood	0.258	Path density	-0.508
Trust propensity	0.233	Average connected distance	0.483
Interest entertainment	0.147	Revisits	-0.383
Interest business	0.254	Compactness	1.500
		Divergent navigation	1.189

Table 4 Predictions of task outcomes based on navigation metrics

Task outcome	R^2	Predictors	Beta
Performance	0.299	Path adequacy	0.385
		Reading time	0.298
Perceived disorientation	0.213	Flimsy navigation	0.462

Table 5 Predictions of task outcomes based on hypothetic factors and navigation metrics

Task outcome	R^2	Predictors	Beta
Performance	0.515	Spatial ability	0.399
		Finance expertise	0.340
		Path adequacy	0.216
		Reading time	0.318
Satisfaction	0.793	Motivation	0.754
		Usability	0.619
		Interest business	-0.277
		View large	0.491
		View small	-0.302
Perceived Disorientation	0.57	Usability	-0.496
		Cycles	-0.388
		Working Memory	-0.305

4 Conclusion, discussion and further research

4.1 Summary of results

This paper has shown that user characteristics such as *domain expertise*, *spatial ability*, *working memory*, *motivation*, and *interest* are important determinants of task outcomes. Interface and context factors such as *sites' usability* and *time constraints* have also an influence on some of the task outcomes.

However, user characteristics as determinants of task outcomes can only be measured in experimental settings. The paper has also shown that some of the user characteristics such as *Internet expertise*, *spatial ability*, *working memory*, *episodic memory*, *trust propensity*, and *interests* can be estimated with a reasonable level of accuracy based on web logging data that can unobtrusively be collected in a real-world navigation session.

The predictions of task outcomes based on user characteristics, interface and context factors appeared to be more accurate than those based on navigation metrics. This difference suggests that there is still enough work to be done in searching for accurate and relevant indicators of navigation behavior. However, both categories of predictors are important, one from a more theoretical perspective and the other from an applied one.

A considerable number of factors proved to be less relevant than expected or reported in literature. For example, the demographic factors gender and age did not correlate with any of the task outcome or navigation metrics. The affective and conative factors were not as important as the cognitive factors.

4.2 Discussion and implications

Studying a large number of factors in relation to a comprehensive range of outcomes of web navigation

tasks in a particular domain (Web-assisted Personal Finance) was useful in several respects. A limited number of significant predictors was identified, and their relative contribution to the accuracy of predictions was estimated. Since factors were studied together and the stepwise method of regression analysis was employed, it was possible to rule out factors that were only marginally significant or confounded with one another. This is an important contribution of the research reported in this paper in comparison with other work of this type. Most of the studies addressing individual differences in web navigation (including those referred in this paper) are restricted to a limited number of user characteristics, and for this reason they can easily overlook other (more important) characteristics. For example, the influence of working memory on hypertext navigation as reported by Tucker and Warr [50] might not have appeared as significant if spatial ability was included as a predictor in their model.

The proposed approach to calculating different types of metrics based on navigation data was proved to be profitable. Different types of knowledge about the user can be inferred based on the kind of information that is extracted from this data: *syntactic* (structural) information indicated mainly users' navigation styles, for example, if they rather revisit pages than viewing new pages, if they return using the back button or just by following links, etc. (first- and second-order metrics); and *semantic* information indicated if users were effective in pursuing their goals (path adequacy) independent of their navigation styles.

Obviously, there are no strong grounds for a large generalization of the obtained results. They apply to situations where goal-directed and performance-oriented tasks are performed. Moreover, the results are valid in the *Desktop* paradigm. Under a *Mobile* paradigm with different site structures and navigation support available, *Flimsy navigation* style, for instance, might not necessarily be associated with *disorientation*. The navigation sessions were restricted to 40 min, which might have prevented some of the users to form an adequate mental model of the web sites used. The number of subjects (30) was rather limited and relatively homogeneous, as they were students. This might have had an impact especially on factor analysis results. The Web-assisted Personal Finance domain might appear as limited although it is frequently mentioned in user surveys [22] and in visible expansion especially for the investigated user population (e.g., e-banking). Restriction of the study to one particular domain was an essential methodological requirement. Crossing multiple domains might have canceled out some of the observed correlations, and in particular the correlation between domain knowledge and task performance. Despite these limitations, there are enough confirmations of previous research [5, 7, 11, 44, 46, 47] to support the reliability of the results.

The results of this study have important practical implications. A web application can be designed in such

a way that it takes into consideration (or compensates for) those factors that proved to be significant in predicting task outcomes. For example, since *spatial ability* is one of the determinants of *task performance*, some interface features (e.g., maps) should be designed to compensate for low spatial ability. The indicators of navigation behavior that are automatically calculated during a navigation session and are able to predict relevant user characteristics and also task outcomes can be used to model the user in real time and personalize the application. For example, the application can be programmed to provide additional navigation aid when users are diagnosed 'at risk of disorientation' and to hide useless hints when users are assessed as 'doing well'.

4.3 Further research

From a theoretical perspective, it appears that spatial-semantic cognitive mechanisms are crucial in adequately performing web navigation tasks. This study has only identified some individual differences that are consistently associated with specific task outcomes. Further research is planned to investigate and to model cognitive mechanisms that are responsible for these individual differences and for their influence on task outcomes. Presently, the existing cognitive models of web navigation [12, 16] ignore almost completely the spatial dimension, and treat solely the semantic dimension of web navigation (information scent).

Only syntactic and semantic metrics were used in this study; extracting from navigation data pragmatic information indicative of users needs, interests, goals and tasks is also considered.

Acknowledgements The study presented in this paper was conducted in collaboration with Eelco Herder and Betsy van Dijk from Twente University, The Netherlands. We would like to thank the anonymous reviewers for their helpful comments on earlier versions of this paper.

References

1. Lazar J (2003) The World Wide Web. In: Jacko J, Sears A (eds) The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. Lawrence Erlbaum, Mahwah, pp 714–730
2. Ivory M, Hearst M (2001) State of the art in automating usability evaluation of user interfaces. *ACM Comput Surveys* 33(4):1–47
3. Conklin J (1987) Hypertext: an introduction and survey. *IEEE Comput Mag* 20(9):17–41
4. Eveland WP Jr, Dunwoody S (1998) Users and navigation patterns of a science World Wide Web site for the public. *Public Understand Sci* 7:285–311
5. MacGregor S (1999) Hypermedia navigation profiles: cognitive characteristics and information processing strategies. *J Educ Comput Res* 20(2):189–206
6. Draper SW (1999) Supporting use, learning, and education. *J comput documen* 23(2):19–24
7. Chen C (2001) Individual differences in a spatial-semantic virtual environment. *J Am Soc Inform Sci* 51(6):529–542

8. Stanney KM, Salvendy G (1995) Information visualization; assisting low spatial individuals with information access tasks through the use of visual mediators. *Ergonomics* 38(6):1184–1198
9. Westerman SJ (1995) Computerized information retrieval: individual differences in the use of spatial vs nonspatial navigational information. *Perceptual Motor Skills* 81:771–786
10. Sjolinder M (1998) Individual differences in spatial cognition and hypermedia navigation. In: *Towards a framework for design and evaluation of navigation in electronic spaces*. Swedish Institute of Computer Science
11. Neerincx MA, Pemberton S, Lindenberg J (1999) U-WISH Web usability: methods, guidelines and support interfaces. TNO Human Factors Research Institute, Soesterberg
12. Pirolli P, Fu WT (2003) SNIF-ACT: a model of information foraging on the World Wide Web <http://citeseer.ist.psu.edu/pirolli03snifact.html>
13. Byrne MD (2003) Cognitive Architecture. In: Jacko J, Sears A (eds) *The human-computer interaction handbook: Fundamentals, evolving technologies and emerging applications*. Lawrence Erlbaum, Mahwah, pp 97–117
14. Card SK, Pirolli P, Van Der Wege M, Morrison JB, Reeder RW, Schraedley PK, Boshart J (2001) Information scent as a driver of web behavior graphs: results of a protocol analysis method for web usability. SIGCHI'01. ACM Seattle
15. Kieras DE, Meyer DE (1997) An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Hum Comput Interact* 12:391–438
16. Kitajima M, Blackmon MH, Polson PG (2000) A comprehension-based model of web navigation and its application to web usability analysis, in people and computers XIV. Springer, Berlin Heidelberg New York, pp 357–373
17. Landauer TK, Foltz PW, Laham D (1998) Introduction to latent semantic analysis. *Discourse Process* 25:259–284
18. Blackmon MH et al (2002) Cognitive walkthrough for the Web. In: CHI 2002. ACM, Minneapolis
19. Jameson A (2003) Adaptive interfaces and agents. In: Jacko J, Sears A (eds) *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*. Lawrence Erlbaum, Mahwah, pp 305–330
20. van Oostendorp H, Goldman SR (eds) (1999) *The construction of mental representations during reading*. Lawrence Erlbaum Associates Publishers, Mahwah
21. van Dijk TA, Kintsch W (1983) *Strategies of discourse comprehension*. Academic, New York
22. Morrison JB, Pirolli P, Card SK (2000) A taxonomic analysis of what World Wide Web activities significantly impact people's decisions and actions. UIR Technical report UIR-R-2000–17, Xerox PARC
23. Pirolli P, Card SK (1999) Information foraging. *Psychological Rev* 106(4):643–675
25. ISO 9241 - 11 Ergonomic requirements for office work with VDT's—guidance on usability
25. Reason J (1990) *Human error*. Cambridge university press, Cambridge
26. van Oostendorp H, Walbeehm BJ (1995) Towards modelling exploratory learning in the context of direct manipulation interfaces. *Interact Comput* 7(1):3–24
27. Herder E, Van Dijk B (2004) Site structure and user navigation: models, measures and methods. In: Chen SY, Magoulas GD (eds) *Adaptable and Adaptive Hypermedia Systems*. Idea Group Publishing, pp 19–34
28. Botafogo RA, Rivlin E, Shneiderman B (1992) Structural analysis of hypertexts: identifying hierarchies and useful metrics. *ACM Trans Inform Syst* 10(2):142–180
29. Hilgard ER (1980) The trilogy of mind: Cognition, affection, and conation. *J Histo Behav Sci* 16:107–117
30. Atman KS (1982) *The conation phenomenon*. University of Pittsburgh School of Education, Pittsburgh
31. Hershberger WAE (1989) Volitional action: conation and control. *Advances in Psychology*, vol 62. Elsevier Science, Amsterdam
32. Wechsler D (1950) Cognitive, conative, and non-intellective intelligence. *Am Psychol* 5:78–83
33. Emmons R (1986) Personal strivings: an approach to personality and subjective well-being. *J Personal Soc Psychol* 51:1058–1068
34. Mischel W (1996) From good intentions to willpower. In: Gollwitzer P, Bargh J (eds) *The psychology of action*. Guilford press, New York, pp 197–218
35. Huiitt W (1999) Conation as an important factor of mind. *Educational Psychology Interactive*. Valdosta State University, Valdosta
36. Linderholm T (2002) Predictive inference generation as a function of working memory capacity and causal text constraints. *Discour Process* 34(3):259–280
37. Pettijohn T (2002) Locus of control scale. Web Survey
38. Peak Performance, Brain Dominance Questionnaire (2002) <http://www.glencoe.com/ps/peak/selfassess/integrate/>
39. Egger FN (2003) From interactions to transactions: designing the trust experience for business-to-consumer electronic commerce. Technische Universiteit Eindhoven, Eindhoven
40. Miron EAM (2003) Analyzing the role of students' goal orientation during their interaction with an ILE. In: HCT2003—the 7th human centred technology postgraduate workshop. HCT, Falmer, Brighton
41. Compeau DR, Higgins CA (1995) Computer self-efficacy: Development of a measure and initial test. *MIS Quart* 19(2):189
42. Wammi, Web Usability Questionnaire. <http://www.wammi.com/>
43. Sumi, The Software Usability Measurement Inventory. <http://www.ucc.ie/hfrg/questionnaires/sumi/>
44. Ahuja JS, Webster J (2001) Perceived disorientation: an examination of a new measure to assess web design effectiveness. *Interact Comput* 14(1):15–29
45. Cohen J (19992) A power primer. *Psycholo Bull* 112(1):155–159
46. Cockburn A, McKenzie B (2001) What do web users do? An empirical analysis of web use. *Int J Hum Comput Stud* 54(6):903–922
47. Rouet JF, Ros C, Jegou G, Metta S (2003) Locating relevant categories in web menus: effects of menu structure, aging and task complexity, in human-centred computing. Cognitive, Social and Ergonomic Aspects. In: Harris D, Duffy V, Smith M, Stephanidis C (eds) *HCI International 2003—Conference Proceedings*. Crete, Greece, Lawrence Erlbaum Associates, Publishers, Mahwah, pp 547–551
48. Juvina I, van Oostendorp H (2003) Human Factors in web-assisted personal finance. Human-centred computing. Cognitive, social and ergonomic aspects. In: Harris D, Duffy V, Smith M, Stephanidis C (eds) *HCI International 2003—Conference Proceedings*. Crete, Greece. Lawrence Erlbaum Associates Publishers, Mahwah, 3:477–481
49. Takeishi A (2001) Bridging inter- and intra-firm boundaries: management of supplier involvement in automobile product development. *Strategic Manage J* 22(5):03–433
50. Tucker P, Warr P (1996) Intelligence, elementary cognitive components, and cognitive styles as predictors of complex task performance. *Personal Individual Differences* 21(1):91–102