A WWW Information Seeking Process Model

Susanne Loeber and Alexandra Cristea
Technische Universiteit Eindhoven
P.O.Box 513, 5600 MB Eindhoven, The Netherlands
s.g.loeber@tue.nl
a.i.cristea@tue.nl

Abstract

In now-a-days information society, most people use the Web for finding information. The correct modelling and interpretation of this information seeking process (ISP) is an extremely hot topic among web designers.

We aim at investigating, analysing and modelling the visitor and the website. These aims are towards establishing a framework of guidelines for web design to enhance visitor goal or task fulfilment and thus visitor satisfaction. On the one hand we show how ISP works and on the other hand we should how it can be used to design a context-driven, rhetorical role-playing approach.

We conclude that the ISP of each individual visitor can be placed in a context (a website where information can be found) and that the interaction between this visitor and website should be show through author persona and user persona.

Keywords

Information-Seeking Process model, Persona theory, MAO model, Navigation

Introduction

Seeking on the World Wide Web (WWW) is not always easy. Visitors often cope with difficulties, such as lost-in-space syndrome, cognitive overhead, closure and/or distraction. When designing a website it is not only important to know how the visitors proceed in gathering information, but it is also important, why they seek information.

We are interested in website design, with special focus on investigating ways to make websites more visitor-friendly for each individual who is part of the target audience. We divide our research goal into five major sub-goals:
1. to show of way of describing websites,
2. to specify visitor information that characterises the Information Seeking Process (ISP),
3. to model the navigation of visitors on a website,
4. to extract website design guidelines,
5. to find ways to evaluate visitor behaviour on different types of websites.

This paper embarks on an analysis in support of the first four sub-goals. In the second section, we show a way of analysing the type of website. Then we describe the navigational behaviour of the visitor (what) and user characteristics (why). By combing this information we built our extended Information Seeking Process (ISP) model. In section three, we show our extended ISP model. With this model we give a more precise description how the visitor seeks for information on websites. This model will be used to extract website design guidelines. In the fourth section, a context-driven, rhetorical role-playing approach to ISP modelling is presented. In this section, we will first describe the Persona Theory and Situated Action framework and then our guidelines. In section five, we finalize with future work, discussions and conclusions.

Gathering information about websites

When gathering important features about websites, we make a distinction between different types of websites. In this section we will try to abstract to a high but workable level. We will focus on area of usage, purpose, value, social navigational tools and target audience.

Websites are being designed for different usage areas: (Brusilovsky, 1996)
- educational hypermedia systems,
- online information systems,
online help systems,
information retrieval systems,
institutional hypermedia systems,
personalized views in information space.

The reason why we have chosen *area of usage* is due to the fact that each area has its own characteristics. This item says something about the target audience (who they are and what their plans are), the known or expected differences between the visitors (different abilities, for example knowledge and opportunities, for example time,) the size of hyperspace, task of the target audience (search vs. work), the goal of the designer on adaptive navigation support, the knowledge about layout, discourse, structure and content selection.

The *purpose* of the website is to inform, to teach, to persuade or to entertain. *To inform* is to formally or officially tell someone about something; to show plain facts without any educational, persuasive, entertaining elements. *To teach* means to show someone how to do something and/or to show or tell someone how they should behave or what they should think; to show courseware. *To persuade* is to make someone decide to do something, to buy something, to adjust to a certain lifestyle, to believe something, etc.; to show facts and fiction is a context with persuasive elements. *To entertain* means to do something that amuses or interests people; to show facts and/or fiction in a context.

The purpose of the presentation says something about the content selection and overall look and feel. Therefore, the reasons to choose purpose as defining item are on one hand, a presentation should have one goal and on the other hand each area of usage can (more or less) the same purpose.

The *value* of a website can be utilitarian or hedonistic. *Utilitarian* means formal useful and practical rather than being used for decoration (e.g., the layout of a search engine). *Hedonistic* means playful, relaxing. A hedonist is someone who believes that pleasure is the most important thing in life. Hedonistic means showing information in such a way that the visitor loves the look and feel (He didn't like the paintings of Rembrandt, but he loved the presentation.) or the tasks he can do (He didn't like the paintings of Rembrandt, but he enjoyed the task of counting all the blond young ladies.).

The value of the presentation says something about the layout and the structure. The reasons to choose this division are simple. Content can be presented something in two ways. First way is by showing them texts and pictures. The second way is by showing them the same information through a game, animation, quiz, etc. The same goes when the purpose of the presentation is to persuade visitors. Unfortunately, this is not the case when the goal is to inform or entertain visitors. There can be made a link between utilitarian value and *do-ers* and between hedonistic value and *viewers or readers*.

The website can have *social navigational tools*. This is outside the scope of this article. For more information, see articles by Dieberger et al. (2001) and Erickson and Kellogg (2000).

The *target audience* of the website can be any group of people. It can exist of anonymous individuals or well-known clients and anything in between.

**Gathering information about visitors**

If we want to know something about the visitors who visit a website, two questions need to be answered. The first question: *What do visitors do on a website?* The second question: *Why do visitors navigate this way?* The first question can be answered when looking at the navigational behaviour and by looking at the way people seek for information. Both answers will be given in the first subsection. In the second subsection the user characteristics will be described. We make a distinction between visitors and users. Visitors are those individuals who visit a website to fulfil a need. They are opposite from users who are made up by the designer. There should be a match between a user (profile) and a visitor. If the match is correct than the visitor is part of the target audience. Each time a visitor visits the website she can be matched to another user, because she has another plan (need, goal, task).
What do visitors do on a website?

How visitors navigate?

When visitors search and/or browse on a website, they navigate through hyperspace in order to find information, to read text, to communicate with other visitors, to play games, etc. We describe searching as typing keywords in order to find information and browsing as following links to get to the information needed. Carmel (1992) distinguishes three types of browsing:

- **search oriented browsing**: the process of scanning to find and review, to integrate information relevant to a fixed task.
- **review browsing**: the process of scanning to find and reviewing, to integrate interesting information in the presence of transient browse goals that are constrained by motivation or interest.
- **scan-browsing**: the process of scanning to find (and not reviewing to integrate) information in the presence of transient browse goals that are constrained by motivation or interest.

Each type of browsing asks for a different structure and content selection. This is also the case when a distinction has been made between visitors who search and visitors who browse to find their information. Browsing can be done with or without the help of other visitors. This is part of social navigation is outside the scope of this article.

![Figure 1. Choo’s ISP Model](image)

**How do visitors find information?**

There are two ways to find information, by *internal search* and *external search*. Internal search is the process of recalling stored information from memory. If the result is not satisfying, then people will start an external search. External search is the process of collecting information from outside sources. These sources, as described by Hoyer and McInnis (2001), can be either online or offline. Some examples are: retailer search (visiting websites of stores), media search (view online adds and other types of marketer-produced communication), interpersonal search (use email, chatbox, FAQ’s to get online advice from friends, staff, and other consumers), independent
search (visit websites with independent sources of information, such as government information, books or magazines) or experimental search (try online a product or software sample, service trails).

When people are motivated to seek for information, the external search will be more extensive. There are six factors that increase the motivation to do an external search: involvement and perceived risk, the perceived costs and benefits of search, the nature of the consideration set (number or alternatives), relative brand uncertainty, attitudes towards search, and the level of discrepancy of the new information. External search is also strongly influenced by people's ability to process information. It is also dependent on the opportunity to process information. Situational factors that might affect the process include the amount of information, the information format and the time available. (Hoyer and McInnis, 2001)

People don't just start and an external (or internal) search. Something has happened that makes them seek for information. It could be the case that she has a problem, is uncertain, has found an ambiguity or is curious about something. To fulfil this need for information, she enters a webpage and starts seeking for relevant information. Choo (1999) describes the external search as being part of an information seeking process (ISP). See Fig. 1. This process contains three steps:
1. information need
2. information seeking
3. information use

Why do visitors navigate this way?

User characteristics

Each visitor differs from other visitors because she has certain characteristics and preferences. There are different views on what features are relevant to interpret, to better react to each visitor's actual needs. Brusilovsky (1996) describes five features used by adaptive hypermedia systems: user's goal, knowledge, background, hyperspace experience and preferences. De Troyer (1998) divides visitors by geographical, demographical and psycho-graphical criteria. Turk (2000) describes visitor characteristics according to usability evaluation: age, gender, culture, disabilities, educational level, WWW/IT experience and interest/expertise. Zeldman (2001) groups websites' audience based on needs and goals into: do-ers (people using tools to accomplish task(s)), readers (people who seek entertainment/relaxation) and viewers (people who regard a web site as a book).

We would stress the importance of navigation. It should be part of the user characteristics, either as known feature or linked to cognitive style or plan. With navigation we mean browsing and searching strategies. These strategies are not included into the WWW/IT experience by Turk (2000). Most of the characteristics mentioned above can be combined inside a Motivation, Ability and Opportunity (MAO) model by Hoyer and McInnis (2001) and will be used inside our ISP model.

<table>
<thead>
<tr>
<th>Motivation: inner state of arousal; energy for goal achieving; factors are:</th>
</tr>
</thead>
<tbody>
<tr>
<td>personal relevance,</td>
</tr>
<tr>
<td>values,</td>
</tr>
<tr>
<td>goal,</td>
</tr>
<tr>
<td>functional, symbolic or hedonistic need of a social or non-social nature,</td>
</tr>
<tr>
<td>perceived risk,</td>
</tr>
<tr>
<td>inconsistency with prior attitudes.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ability: extent of consumers' action resources; factors are:</th>
</tr>
</thead>
<tbody>
<tr>
<td>knowledge and experience,</td>
</tr>
<tr>
<td>cognitive style,</td>
</tr>
<tr>
<td>intelligence, education, age,</td>
</tr>
<tr>
<td>monetary resources.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Opportunity: acting chance when easier (than usual); factors are:</th>
</tr>
</thead>
<tbody>
<tr>
<td>time,</td>
</tr>
<tr>
<td>distraction,</td>
</tr>
<tr>
<td>information amount,</td>
</tr>
<tr>
<td>complexity,</td>
</tr>
<tr>
<td>repetition. (local minimum on constraints surface).</td>
</tr>
</tbody>
</table>

Table 1. MAO model
Hoyer and McInnis (2001) place the need for information inside their *Motivation, Ability and Opportunity* (MAO) model (Table 1). They explain that MAO is relevant to a wide variety of visitor behaviours, e.g. navigation on websites. Motivation, Ability and Opportunity can be influenced by the factors as described in Table 1. It is important when describing the navigational behaviour of the visitor on a website to take these factors into account.

**Our ISP model**

We have extended the ISP model of Choo with MAO of Hoyer and McInnis. It still contains three steps and will be described in the first subsection. In the second subsection we provide a pseudo-mathematical expression of our model.

**The three steps of our ISP model**

Our extended version of the information-seeking process model must be particularly suited to navigation on the WWW. We have identified three steps, see Table 2, by integrating the general information seeking model above with the needs model and the MAO model, as follows:

1. the visitor experiences a (non-)social need in the context of a/some
   a. functional need(s) and/or
   b. symbolic need(s) and/or
   c. hedonistic need(s).
2. the visitor seeks for information on the WWW
   a. to find information about X.
   b. to find a way to enhance/change their self-image, public image or esteem.
   c. to have sensory pleasure.
3. the visitor uses the information found
   a. to read, buy, compare, verify, exchange, or … X.
   b. to interact with others, copying others behaviour etc,
   c. to have pleasure.

<table>
<thead>
<tr>
<th>Table 2. Our ISP model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. the visitor experiences a (non-)social need in the context of a/some</td>
</tr>
<tr>
<td>a. functional need(s) and/or</td>
</tr>
<tr>
<td>b. symbolic need(s) and/or</td>
</tr>
<tr>
<td>c. hedonistic need(s).</td>
</tr>
<tr>
<td>2. the visitor seeks for information on the WWW</td>
</tr>
<tr>
<td>a. to find information about X.</td>
</tr>
<tr>
<td>b. to find a way to enhance/change their self-image, public image or esteem.</td>
</tr>
<tr>
<td>c. to have sensory pleasure.</td>
</tr>
<tr>
<td>3. the visitor uses the information found</td>
</tr>
<tr>
<td>a. to read, buy, compare, verify, exchange, or … X.</td>
</tr>
<tr>
<td>b. to interact with others, copying others behaviour etc,</td>
</tr>
<tr>
<td>c. to have pleasure.</td>
</tr>
</tbody>
</table>

Step one starts with a sensory or mental input that triggers the visitor’s attention. If the visitor decides not to ignore it and is not able to solve it by performing an internal search, then it will transform into an information need plus a functional, symbolic or hedonistic need of social or non-social nature. When the visitor feels such need(s), she experiences an inner state of tension caused by a (perceived) distance from an ideal or desired physical or mental state. The visitor attempts to regain a sense of equilibrium by conceptualising some goal supported by task(s) in order to make their information needs clear and feasible and fulfil them.

During the second step, the visitor will seek information on the WWW. This typically consist of looking for texts, pictures, videos, sounds, etc. for which the visitor browses or searches on the website.

Step three starts when the visitor has completed the browsing and/or search task(s). In the ideal case, the visitor has found what she was looking for. Once accessed, the information can be used to satisfy the perceived need. For example, the visitor reads an article on health issues to stay well informed (information + functional needs), buys a bigger car to show off when meeting peers (information + symbolic needs) or tries a recipe for that delicious looking cake (information + hedonistic needs).

Our model is illustrated in Figure 2. This model attempts to provide the researcher at least with the possibility to describe the errors made by the visitor (Sutcliffe, 1999) by adding error modelling and to express MAO features. This model is built of different navigational models, see Loeber and Cristea (2002). The motivation, opportunity and maybe even the ability of the visitor can change during the visit on the website. This can also be shown when describing the MAO part of the model in each of the steps.
The pseudo-mathematical expression of our ISP model

It is important not only to design models, but to make them also computable. A pseudo-mathematical expression of our Internet navigation model is presented in the following. (See Table: 3)

Although the model is able to show why and in what way the visitor does an external search, it is still impossible to check or evaluate this. This could be solved by building a MAO layer inside a presentation generation engine and perform consistency checks. How this can be done is described in Loeber et al. (2002). The next step is to use our model to extract website design guidelines. These guidelines will help to focus on both questions from the perspective of the visitor and the company, institute, etc. who offers the online information.
How to use our extended ISP model

First we will describe the Persona Theory and Situated Action framework, because these are the pillars of our context-driven, rhetorical-role-playing approach to ISP. Next, we will combine the theory with the framework and then we will extract website design guidelines.

The Persona Theory

The starting point of this theory is: the premises that people communicate through a series of rhetorical roles they assume appropriate to fulfil their needs and that are correct within the current situation. Persona Theory (Coney and Steehouder, 2000) focus on rhetorical role-playing of users based on their rhetorical purposes within texts. A rhetorical purpose is a plan that consists of a social or non-social need of a functional, symbolic, or hedonic nature, a goal and tasks/decision strategies. A rhetorical role is the means by which the implied author and the implied user communicate with each other. In order to play a role as implied author or implied user, two personas are needed: an author persona and an user persona.

The role of the implied author or author persona is ‘created’ by the content and style of the web site; the voice (e.g., of the head of the company or owner of the web site) that speaks through the web site. Typical rhetorical roles of the author persona on a web site of a university could be: Head of University, Teacher, Parent from the parent-board, First-years-student Advisor, etc.

To be able to design a credible and inviting author persona, the designer should be kept in mind when taking into consideration the presence of the author persona. (Key decisions are: should the author persona be anonymous or identifiable? Should the author persona have a strong presence? Should peripheral cues be used to reinforce the authors persona?); taking into consideration the role(s) of the author persona. (Key decisions are: Which role(s) will the author persona take? Will the author persona take one or more roles? How will the role(s) be prompted?) and taking into consideration the values of the author persona(s). (Key decisions: Which of the organization’s values should be emphasized? How explicitly should these values be presented?) (Coney and Steehouder, 2000)

The role of the implied user of user persona is ‘created’ by the content and style of the web site. Typical rhetorical roles of the user persona on a web site of a university are: student, (grand) parent of participating student, teacher (from another country), manager from a company, etc.

To be able to design playable and attractive user personas, the MAO model can be kept in mind when defining the role(s) of the user persona(s). (Key decisions are: What role(s) should the implied user or user persona have? Which goals (and needs) should the user persona have? How should the user persona(s) be prompted?) and

Table 3. Pseudo-mathematical expression of our ISP model

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>prior knowledge(A) # A is the visitor’s prior knowledge</td>
</tr>
<tr>
<td>2.</td>
<td>¬ know(X) # need of X</td>
</tr>
<tr>
<td>3.</td>
<td>know(Y) ⇒ know(X) # knowing Y implies knowing X</td>
</tr>
<tr>
<td>4.</td>
<td>X ⊆ Y # Y is a reformulation of X in context; this makes X searchable on the WWW</td>
</tr>
<tr>
<td>5.</td>
<td>case 1: Y ⊆ A =&gt; END # if Y was contained in initial prior knowledge, X can be deducted, so no search is necessary</td>
</tr>
<tr>
<td>6.</td>
<td>case 2: Y ⊄ A =&gt; SEARCH(Y, WWW) # search for Y on WWW</td>
</tr>
<tr>
<td>7.</td>
<td>subgoal(Y) # Y becomes the next subgoal (if prior goal already exists)</td>
</tr>
<tr>
<td>8.</td>
<td>select_strategy(Y)</td>
</tr>
<tr>
<td>a.</td>
<td># a search strategy should be selected, e.g., search, browse, chaining, monitoring, etc.; or combination of the above</td>
</tr>
<tr>
<td>b.</td>
<td>quality criteria should be selected (i.e., how good the solution should be)</td>
</tr>
<tr>
<td>c.</td>
<td>quantitative criteria should be selected (i.e., how many solutions are needed)</td>
</tr>
<tr>
<td>d.</td>
<td>end (stop) criteria should be selected (timeout, number of steps, etc.)</td>
</tr>
<tr>
<td>9.</td>
<td>run_search # this is the “aquire data” phase; search is performed conform with decision strategies</td>
</tr>
<tr>
<td>10.</td>
<td>interpret_effect # here, the solution can be evaluated, diagnosed, error correction performed</td>
</tr>
<tr>
<td>11.</td>
<td>use(X) # this is the information use stage</td>
</tr>
<tr>
<td>12.</td>
<td>interpret_effect # here, the usage as well as the solution can be evaluated, diagnosed, error correction performed</td>
</tr>
</tbody>
</table>

Table 3. Pseudo-mathematical expression of our ISP model
considering the values to be attributed to the user persona. (Key decision: Should values be expressed directly or indirectly?) (Coney and Steehouder, 2000).

Persona Theory does not view the user as a passive reader, but as participant actively engaged in a process of making meaning, cooperating closely with the author and playing a certain role in order to fulfil her need. From this point of view, the concept of role-playing serves as a very insightful and useful design tool when designing a web-based presentation, as opposed to simple stereotypes of the type proposed by Elaine Rich (1979). Persona Theory makes it possible to design a context in which multiple user personas and multiple visitors who are matched to one and the same user persona can interact with each other.

The Situated Action framework

Situated Action underscores the view that every course of action depends in essential ways upon its material and social circumstances (Suchman, 1987). Their focus is towards the situation in which the activity of the visitors interacting (role-playing) takes place (is situated: situatedness). They are interested in understanding how and to what extent the situation influences the behaviour of the visitor when the situation is used as resources. We define, based on Lueg’s (2001) definitions, “situation” as physical and social environment in which situatedness—using the situation as resource for action—is happening and “context” or “setting” as an abstraction of a situation in the sense that particular aspects of the situation have been isolated and can be manipulated. A context is for example an author persona and an user persona who interact in such a way that certain content, the presentation format (layout, discourse style, social navigational tools and interaction level) and certain medium used to interact are shown to the visitor.

By combining the Persona Theory with the Situated Action framework, a context-driven rhetorical role-playing approach arises. The power of our approach is: the possibility to predict more precisely and understand more clearly the behaviour of the visitor and at the same time focus on her preferences. This is due to the context aspects and what drives the user to behave in this way (or close to the way, the implied user would behave).

Website design guidelines

These guidelines have a context-driven, rhetorical role-playing approach and are based on Persona Theory, Situated Action framework and ISP modelling.

To create the user persona a stereotype user must be created. The most important aspect is the plan of the user persona. The plan consists of need, goal and tasks. Also the (dis)abilities of the user persona and the opportunity should be taken into account. To create a playable and attractive user persona the questions as addressed in the section on Persona Theory can be used.

To create a credible and inviting author persona the questions as addressed in the section on Persona Theory can be used. Also the following questions, based on the characteristics of the website, should be answered:
1. What is the area of usage?
2. What is the purpose of the website?
3. What is the value of the presentation?
4. What type of communication will occur between the author persona and the user persona?
5. Who are the favourite user personas?

The interaction between author persona and user persona (question 4) has to do with the level of interaction, discourse style, hierarchy between both personas and the usage of social navigational tools.

The interaction level can be low-medium-high and show the personality of the author. Low means anonymous and high means for example Jan Jansen the Dean of the University. The discourse style has, for the moment, only to do with informal and formal language. The hierarchy between author and user persona has to do with the fact if the author feels more important (has more knowledge, higher in status, etc.) or equal to the user persona. For example the relationship between the Dean and students will be in such a way that the Dean is higher in hierarchy. Whether or not to use social navigational tools, thus whether or not to involve data or interaction from other visitors means whether or not to help the visitor fulfil her social needs.
By naming the favourite user personas (question 5), a link between author persona and user persona emerges. This is the starting point for further adaptation towards the visitor of the website.

**Future work and discussion**

Our model and guidelines are now directed towards hypertext, not so much towards hypermedia/multimedia presentation. We hope in the future concentrate on multimedia to broaden the usage of our model and guidelines. We presume that the author persona description needs to become more complex to streamline the presentation. There should also be a mechanism to match visitor and user persona and at the same time check whether or not the MAO of the visitor changes during her visit. If this is the case a new match has to be found.

A next step will be to alter the look-and-feel of the website not only to the user persona but also directly to preferences and characteristics of the visitor. In this way our guidelines will help to make an adaptive, adaptable website.

We also hope to focus on the different types of communication between author persona and user persona. We believe our discourse style list should be extended with style rules, guidelines and tests to make the design of multiple author personas and user personas more easily. It should also help to link the rules to the (already existing) objects in databases.

Our personas aren’t visualized as ‘puppets’, like for example the Microsoft Paperclip. If we would like to include this type of author persona, then we should make some adjustments to our guidelines. The author persona becomes a split personality; on the one hand is still provides the look-and-feel of the website but on the other hand it also provides it’s own appearance.

As mentioned in step 5, we still need to evaluate our model and guidelines. The testing will need to be done in different areas of usage to understand the power and disadvantages of our ideas.

**Conclusions**

Our agenda has focused on five elements:

- describing the navigation of the visitor on a website,
- describing the ISP of a visitor who navigates on a website,
- describing the behaviour and characteristics of the target audience of a website,
- showing how external search, navigation and characteristics are combined into an extended ISP model,
- showing how our website design guidelines can be extracted from our extended ISP model.

In this paper, we have described the WWW navigation and external search process as sub-categories of a generalized Information Seeking Process (ISP). We have added relevant features to this ISP model and extended it towards a better interpretation of WWW user behaviour. This model in combination with Persona Theory and Situated Action framework is also used to form the base for our context-driven, rhetorical role-playing approach to website design guidelines.

**References**


