User Modeling and Personalization 10: Information Foraging and Web Navigation

Eelco Herder

L3S Research Center / Leibniz University of Hanover
Hannover, Germany

20 June 2016

Outline I

Outline II

Introduction

Hypertext

A Hypertext is a *graph* with uni- or bi-directional *edges*. A *node* represents a text document. An edge between two nodes represents a *link* between two text documents.

Unlike books, hypertext is non-linear. Links allow for associative leaps from one idea to the next; they can connect related Web services; a menu with links provides structure to a hyperdocument or Website.

Orientation and Disorientation

Web navigation is not just about getting from one page to another; it's also about *orientation*.

Adaptive hypermedia are an answer to the problem of disorientation, commonly known as "Lost in Hyperspace"

While navigating a site, users generally need to know:

- identify their current position in a hyperdocument;
- reconstruct the way that led to this position;
- distinguish among different options for moving on from this position.

Knowing the current location

Knowledge about the current location helps users in:

- understanding page content by its relation to other pages
- estimating the level of detail of information provided in a page
- deciding whether one should look further or not

Teevan, J., Alvarado, C., Ackerman, M. S., Karger, D. R. (2004). The perfect search engine is not enough: a study of orienteering behavior in directed search. Proc. CHI 2004.

Theoretical models

The task of user interface designers is to understand the users and their tasks, and how to translate this knowledge into a working system.

Empirical data is a valuable source of knowledge on how users interact with Web-based systems.

Theoretical models, or *cognitive frameworks* explain the underlying processes of observed patterns.

Cognitive frameworks

Cognitive frameworks are working computer programs that simulate user interactions with an interface, making use of strategies that are assumed to reflect actual user strategies. From a behavioristic point of view, these frameworks can be seen as an 'explanation' of the underlying actions.

Information seeking

Information seeking refers to all of the various activities people undergo to find information.

Not all information seeking is the same. Sometimes you may know exactly what you're looking for. Other times, you may have only a general idea of what you're after.

Modes of information seeking

On the Web, several modes of information seeking (or browsing) can be observed:

- *Serendipitous browsing:* there is no real goal and little focus, for example reading through news sites, browsing online stores, checking out Facebook
- *Exploratory seeking:* there may be some idea or a need, but people may not be able to articulate that need clearly. For example, looking for nice holiday destinations.
- *Semi-directed browsing:* the user has a general goal, but no clear idea where to find what he needs. For example, learning activities.

- *Known-item search:* (a.k.a. directed browsing) users know what they want, can describe it in words and may also know where to start looking. For example, browsing Amazon for a specific book.
- *Re-finding:* this often overlooked mode refers to looking for things people have already seen. For example, a known hotel booking site or a blog that they want to share.

In the next part of this lecture, we focus on a model for (semi-)directed browsing.

Foster, A., Ford, N. (2003). Serendipity and information seeking: an empirical study. Journal of Documentation, 59(3), 321-340.

Information Foraging

Developed by Peter Pirolli and Stuart K. Card at PARC in the late 1990s.

A theory that applies the ideas from *optimal foraging theory* to understand how human users search for information.

Assumption: when searching for information, humans use "built-in" foraging mechanisms that evolved to help our animal ancestors *find food*.

The theory assumes that people, when possible, will modify their strategies to *maximize their rate of gaining* valuable information.

Pirolli, P., Card, S. (1999). Information foraging. Psychological review, 106(4), 643.

Overview

Information Foraging Theory consists of three parts:

- *Information patch models*, which deal with time allocation and information filtering activities
- *Information scent models*, which address the identification of information value from proximal cues (content and links)
- *Information diet models*, which address decisions about the selection and pursuit of information items.

Information foraging does not assume that human behavior is classically rational. Decisions are based on:

- imperfect and incomplete information
- balancing effort and expected outcomes (*satisficing*)

Motivation

Imagine a bird of prey that faces the recurrent problem of deciding what to eat.

- different types of environments and prey will yield different amounts of net energy
- the different food-source types will have different distributions over the environment
- this means that different environments or prey will have different access (or navigation) costs
- birds have developed strategies for maximizing the amount of energy returned per amount of effort

Optimal foraging theory models how foragers maximize the rate of gain, given the constraints of the environment in which it lives.

Information seekers (e.g. office workers or researchers) face the recurrent problem of finding task-relevant information.

- information is available in books, manuscripts, online documents, . . .
- different information sources will have different profitabilities (in terms of information returned and time needed to process the source)
- different kinds of sources are distributed in different ways

In typical offices, material for ongoing tasks is readily in hand (in stacks and piles), a personal archive is located in shelves and drawers, other archival information is stored in libraries and other archives.

Information Patches

Optimal foraging theory assumes that the environment of some animal has a ‘patchy’ structure.

- berries are found in patches on berry bushes
- the berry forager must expend some amount of *between-patch* time to get to the next food patch (e.g. berry bush)
- once in a patch, the forager engages in *within-patch* foraging
- as the animal forages within a patch, the amount of food diminishes or depletes
- the animal faces the decision of continuing to forage in the patch or leaving to seek a new one

By analogy, the task environment of an information forager often has a patchy structure:

- information may reside in piles of documents, file drawers, book shelves, libraries, . . .
- online information patches include Web sites, lists of search results, bookmark collections, . . .

Often the person is faced with similar decisions as the berry forager: how should time be allocated among between-patch foraging tasks and within-patch foraging tasks?

Patch Models

Let G be the total net amount of valuable information gained, T_B the amount of time spent between patches, and T_W the amount of time spent within patches.

The rate of gain R can be defined as:

$$R = \frac{G}{T_B + T_W}$$

Further, let t_B be the average time between processing patches and t_W the average time to process a patch.

The average rate of encountering patches is:

$$\lambda = \frac{1}{t_B}$$

The time spent within patches can then be represented as:

$$T_w = \lambda T_B t_w$$

The total amount of information gained can be represented as a linear function of between-patch foraging time and the average gain per item g:

$$G = \lambda T_B g$$

This allows us to express the rate of gain as:

$$R = \frac{\lambda T_B g}{T_B + \lambda T_B t_w} = \frac{\lambda g}{1 + \lambda t_w}$$

The profitability π of a patch is defined as the ratio of net value gained per patch to the cost of within patch processing:

$$\pi = \frac{g}{t_W}$$

Increase of the rate of gain can reached by:

- increasing the profitability π of within-patch activities (e.g. reducing the cost of within-patch processing)
- decreasing the between-patch costs (or rather, increasing the average rate of encountering patches λ

A variation of the above formula for the rate of gain takes into account that there are different kinds of information patches, each with their own rate of encountering λ_i, their amount of information returned in some within-patch foraging time t_{Wi}, $g_i(t_{Wi})$:

$$R = \frac{\sum_{i=1}^{n} \lambda_i g_i(t_{Wi})}{1 + \sum_{i=1}^{n} \lambda_i t_{Wi}}$$

Maximizing the Rate of Gain

When foraging within a patch, the cumulative gain function $g_i(t_{Wi})$ is decelerating. According to Charnov's *Marginal Value Theorem* a forager should remain in a patch so long as the slope of g_i is greater than the average rate of gain R for the environment:

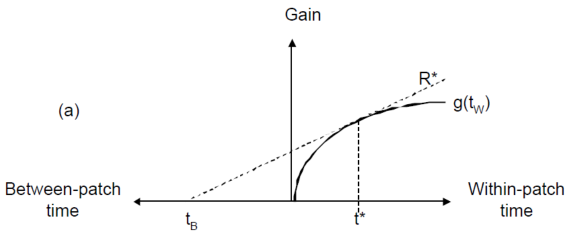


Figure: The rate-maximizing time to spend in a patch occurs when the slope of the within-patch gain is equal to the average rate of gain.

Decreasing time between patches

The figure below illustrates the effect of decreasing between-patch time costs: it does not only improves the overall average rate of gain, but the optimal gain is also achieved by spending less time within a patch.

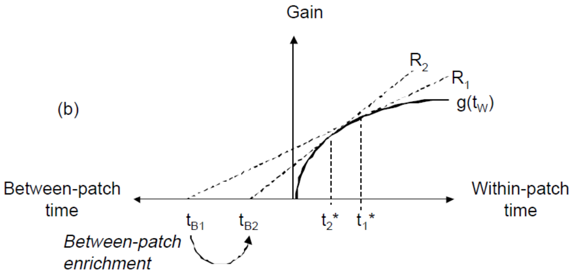


Figure: Enrichment activities that reduce between-patch time costs.

Decreasing time within patches

Within-patch optimization improves the overall average rate of gain as well. A counter-intuitive effect is that the optimal within-patch time is reduced as well.

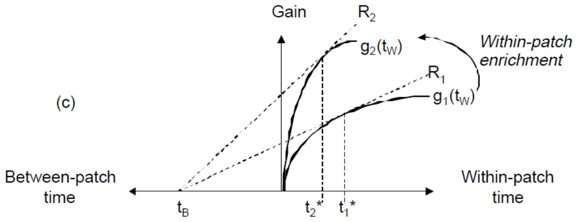


Figure: Enrichment activities that improve the returns from a patch.

Application: Why Google Makes People Leave Your Site Faster

Moving between sites has always been easy.
But, from an information foraging perspective, it used to be best if users stayed put:

- the vast majority of websites were horrible
- the probability that the next site would be any good was extremely low.

https://www.nngroup.com/articles/information-scent/

Early website designers were advised to follow two design strategies:

- convince users that the site is worthy of their attention: having good information and making it easy to find.
- once they arrive, make it easy for users to find even more good stuff so that they stay rather than go elsewhere.

Google has reversed this equation: it is now extremely easy for users to search and find other good sites.

The better search engines get at highlighting quality sites, the less time users will spend on any one site.
Google and always-on connections have changed the most fruitful design strategy to one with three components:

- Support short visits; be a snack
- Encourage users to return; use mechanisms such as newsletters as a reminder
- Emphasize search engine visibility and other ways of increasing frequent visits

Information Scent

As users browse the WWW, they make judgments about the utility of different courses of action available to them. Typically, they must use local cues, such as link images and text, to make navigation decisions.

Information scent refers to the local cues that users process in making such judgments. The analogy is to organisms that use local smell cues to make judgments about where to go next (for instance in pursuing some prey).

Information Scent

the (imperfect) perception of the value, cost or access path of information sources obtained from proximal cues

Application: Deceivingly Strong Information Scent Costs Sales

Users were looking for a baby seat for their car

Quite logically, they looked in the automotive section. No baby seats there, so no sale.

Users assumed that the site didn't sell the product they needed because it wasn't in the category where they assumed they'd find it.

In fact, the product was in a different section of the site, without a cross-reference from the car area. This often ends up costing websites sales.

https://www.nngroup.com/articles/wrong-information-scent-costs-sales/

Information Diet Selection

Diet models in optimal foraging theory deal with the problem of constructing the diet that optimizes its gain of energy per unit cost.

- Pursuing small hard-to-catch prey while large easy-to-catch prey are equally available would be a suboptimal diet.

Remember the formulas for the rate of gain R and the profitability π of an item or patch:

$$R = \frac{\sum_{i=1}^{n} \lambda_i g_i(t_{Wi})}{1 + \sum_{i=1}^{n} \lambda_i t_{Wi}}$$

$$\pi = \frac{g}{t_W}$$

The most logical strategy for information diet selection would be to:

- first order the items by their (estimated) profitability (in decreasing order)
- add items to the diet until the rate of gain for a diet of the top k items is greater than the profitability of the $k+1^{th}$ item.

$$R(k) = \frac{\sum_{i=1}^{k} \lambda_i g_i(t_{Wi})}{1 + \sum_{i=1}^{k} \lambda_i t_{Wi}} > \frac{g_{k+1}}{t_{Wk+1}} = \pi_{k+1}$$

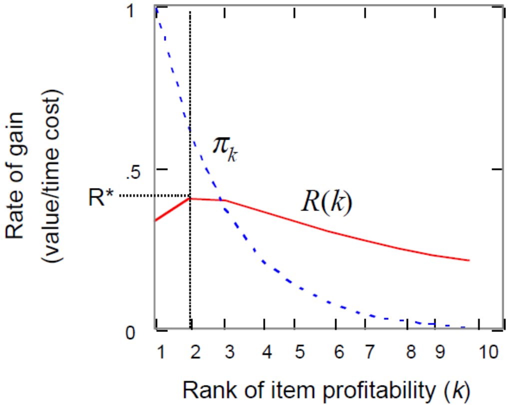


Figure: A hypothetical example of the relationship between profitability (p) and rate of gain (R) for diets including items 1, 2, ...k.

Effects of Scent and Breadth on Use of Search

Katz and Byrne carried out a study with four online stores with menus that differed in:

- breadth (the amount of items displayed at once)
- information scent (the clarity of the labels).

Except for the site title and the color scheme, all other elements were kept constant.

Katz and Byrne. Effects of Scent and Breadth on use of Site-Specific Search on E-Commerce Web Sites. ACM Trans. CHI 10 (3), 2003

Participants were shown each site one at a time in random order. They were asked whether they would click on a menu item or if they would use site search to find the item.

The hypothesis was that ineffective menu structures would lead to higher use of the search functionality.

Which structure performed best?

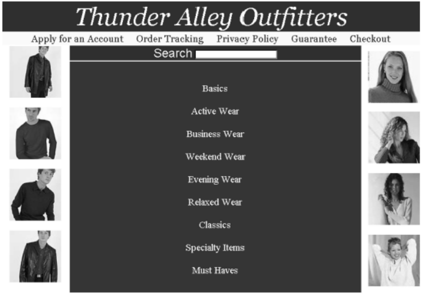
Broad menus, different labels

(a)

(b)

Narrow menus, different labels

(c)



(d)

Results

The results clearly support the notion that the decision to search or to browse is strongly affected by the labeling and menu structure.

Given broad, high-scent menus, participants searched less than 10% of the time, but they searched almost 40% of the time when faced with narrow, low-scent menus.

The breadth effect turned out to be larger than the scent effect (this may not generalize to all sizes of menus and all levels of semantic relatedness).

Social Information Foraging

Information foraging theory has mainly focused on information seeking by the solitary user.
In the Web 2.0 world, social bookmarks and tags turn the discovery of knowledge into a collective action.

Cooperation may yield more benefits than simply making information search more parallel and making it less prone to failure.

- Membership in a group provides resources that can be used to achieve individual goals.

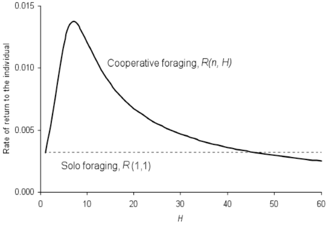
Pirolli, P. (2009). An elementary social information foraging model. Proc. CHI 2009.

Individual foragers who *broker* information and ideas across groups might be exposed to a greater *diversity* of information themselves, and be a conduit to greater diversity for their colleagues.

Effective work groups are ones that *share* information and know-how with external members; effectiveness is improved by *diversity* of the group.

Groups do not become arbitrarily large: there may be some form of *overhead* that at some point outweighs the advantages of further increments in the size of groups.

- limited resources
- group communication costs



The optimum size of this group is n=7. However, people will continue to join the group as long as the rate of return is higher than for solitary foraging.

The rate of gain for the group diminishes, but remaining in the group is still better.

Once the group reaches the *equilibrium size* of n=45, some group members may leave.

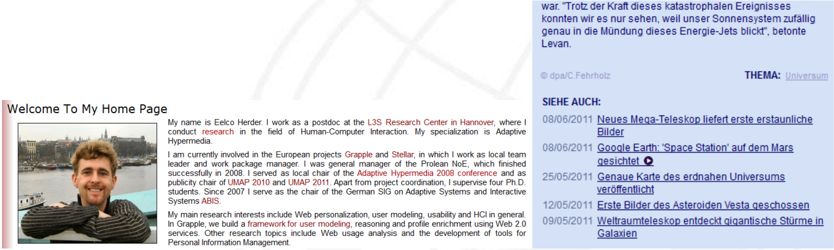
Web Navigation

Information foraging deliberately does not specify what exactly constitutes within-patch and between-patch navigation. On the Web, there is a large number of basic mechanisms that together form the 'navigation logic' on the Web.

In the remainder of this lecture, a survey of the most common mechanisms.

Associative Navigation

Associative links connect pages with similar topics and content. These links may be embedded in running text or displayed as related links.



Welcome To My Home Page

My name is Eelco Herder. I work as a postdoc at the L3S Research Center in Hannover, where I conduct research in the field of Human-Computer Interaction. My specialization is Adaptive Hypermedia.

I am currently involved in the European projects Grapple and Stellar, in which I work as local team leader and work package manager. I was general manager of the Prolean NoE, which finished successfully in 2008. I served as local chair of the Adaptive Hypermedia 2008 conference and as publicity chair of UMAP 2010 and UMAP 2011. Apart from project coordination, I supervise four Ph.D. students. Since 2007 I serve as the chair of the German SIG on Adaptive Systems and Interactive Systems ABIS.

My main research interests include Web personalization, user modeling, usability and HCI in general. In Grapple, we build a framework for user modeling, reasoning and profile enrichment using Web 2.0 services. Other research topics include Web usage analysis and the development of tools for Personal Information Management.

war. "Trotz der Kraft dieses katastrophalen Ereignisses konnten wir es nur sehen, weil unser Sonnensystem zufällig genau in die Mündung dieses Energie-Jets blickt", betonte Levan.

© dpa/C.Fehrholz

THEMA: Universum

SIEHE AUCH:

08/06/2011 Neues Mega-Teleskop liefert erste erstaunliche Bilder

08/06/2011 Google Earth: 'Space Station' auf dem Mars gesichtet

25/05/2011 Genaue Karte des erdnahen Universums veröffentlicht

12/05/2011 Erste Bilder des Asteroiden Vesta geschossen

09/05/2011 Weltraumteleskop entdeckt gigantische Stürme in Galaxien

Figure: Associative and related links

Recommendations and other kinds of adaptive suggested links also fall in the category 'associative links'.

Figure: Recommendations

Local Navigation

Step and page navigation allows people to move sequentially through pages. They provide local structure.

Step navigation typically consists of a text label ('previous', 'next') and an arrow. Step navigation can be regarded as 'direct guidance' through an information domain.

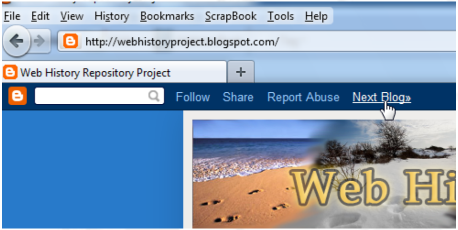


Figure: Step navigation

Page navigation provides additional information and options. Paging is a common method to divide large chunks of information (long texts or search result sets) in smaller pieces.

Figure: Page navigation

Structural Navigation

Structural navigation connects one page to another; they impose a *hierarchy* on the hypertext structure.

The most common and most flexible form of structural navigation is the vertical menu. The menu may be static or it may expand sub-items once a main item is chosen. There are also dynamic variations known as fly-out, pull-down or pop-up menus.

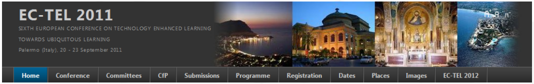


Figure: Navigation bar

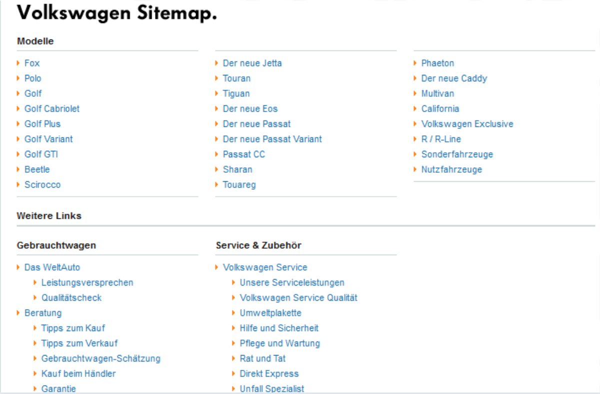
Tab navigation is quite similar to navigation bars, except that each tab stands out individually. Amazon has experimented with many types of tab navigation.

Figure: Tab navigation in Amazon

Overview Navigation

Similar to menus and navigation bars, overview navigation tools impose a (hierarchical) structure upon a Web site.

The most common overview is the site map.



The Open Directory Project DMOZ maintained a (incomplete) directory of the Web. It is currently hosted at Google.

With the increase in quality of search results and the size of the Web, directories have become less popular.

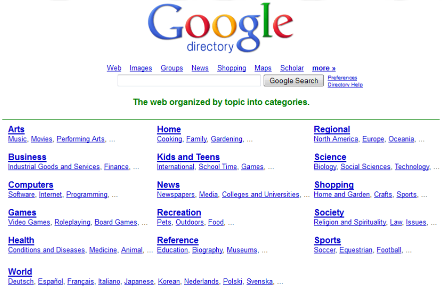


Figure: Web directory

A more recent overview mechanism is the tag cloud, which lists links alphabetically and weighted by frequency or popularity: the larger the link, the more important it is.

Tag clouds reflect a certain *zeitgeist* for a site or topic. As a navigation mechanism, tag clouds have limited value.

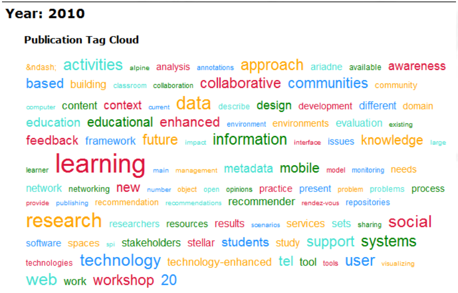


Figure: Tag cloud

Functional Navigation

Functional navigation is a type of meta-navigation that brings the user to specific parts of a site or the Web that are not directly linked, or that are not part of the main content.

The most popular form of functional navigation is the search box.

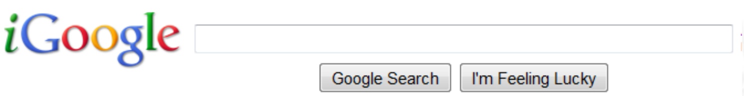


Figure: Google Search

Toolboxes allow users to change their profiles, their settings or (language) preferences.

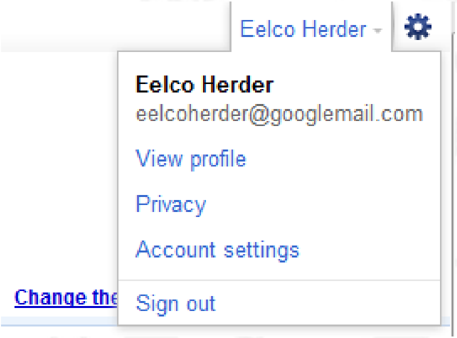


Figure: Toolbox at Google

Submission forms allow people to submit information, to create an account, to provide a transaction or to obtain specific information based on their given input.

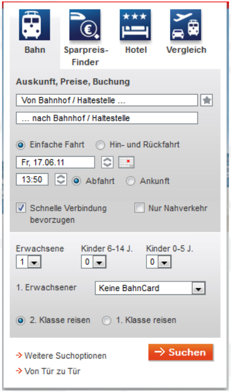


Figure: Submission form at Bahn.de

Temporal Navigation

Temporal (or history) navigation tools allow users to return to pages that they have visited before.

Most temporal navigation tools are integrated in the browser: the back button, bookmarks, tabs and the history list.

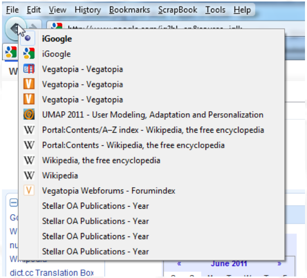


Figure: Browser Back Button

Many sites also offer *breadcrumbs*, links to previously visited pages or (more commonly) pages that are higher in the hierarchy.

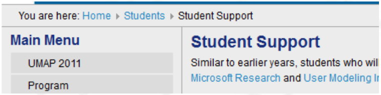


Figure: A breadcrumb bar

Wrapup

While navigating a site, users generally need to know:

- identify their current position in a hyperdocument;
- reconstruct the way that led to this position;
- distinguish among different options for moving on from this position.

On the Web, various different kinds of navigation mechanisms are available to support this process:

- Associative navigation
- Local navigation
- Structural navigation
- Overview navigation
- Functional navigation
- Temporal navigation

Information foraging theory suggests that users:

- look for information and services that is organized in a 'patchy' structure
- make use of proximal cues (information scent) to decide whether to leave a patch or not
- continue searching for information and services until they are satisfied or until the rate of gain drops

Adaptive navigation support aims to address one or more of the above issues by adding, removing, annotating, highlighting or disabling links, by providing personalized overviews or history support.