Metrics for the Adaptation of Site Structure

Eelco Herder
Faculty of Computer Science
University of Twente – The Netherlands
e.herder@cs.utwente.nl

Abstract

This paper presents an overview of metrics for web site structure and user navigation paths. Particular attention will be paid to the question what these metrics really say about a site and its usage, and how they can be applied for adapting navigation support to the mobile context.

Introduction

Device characteristics, such as screen size and means of interaction, and the contexts in which such a device is used, affect a user's navigation through a web site. Past research has revealed that handhelds are typically used in a more goal-directed way than desktop devices [6]. If navigation support does not fit such variable user goals, it will be harder for the users to complete their tasks [10]. Therefore, with the increasing variety of handheld and desktop devices, the need for adapting not only site content, but also site structure to these factors grows [8].

Web site users need *orientation facilities* to identify their current position within a site, to reconstruct the path that led to this point and to distinguish among various options for moving on from this position. Moreover, users also need *navigation facilities* to make their way through the web site [19]. Adaptive navigation support is a reasonable choice for facilitating or even encouraging the most appropriate navigation strategy in a certain usage context [4].

Traditional adaptive hypermedia methods employ user characteristics such as preferences, knowledge and goals. From several research projects it can be concluded that metrics based on site structure and user navigation paths are very useful in predicting web site usability [1], [17]. Moreover, these metrics are expected to be applicable for personalizing navigation support to meet particular objectives or needs of users in different contexts of use [12].

This paper presents an overview of metrics for web site structure and user navigation paths, derived from both adaptive hypermedia research and graph theory. Particular attention will be paid to the question what these metrics really say about a web site and its usage, and how they can be applied for adapting navigation support to different devices and contexts of use.

The Site Graph and the User Navigation Graph

Hypermedia documents can be seen as a collection of interconnected pages. This node-and-link structure is commonly modeled as a directed graph [1], [2], [12] – the *site graph*. User navigation can be seen as an overlay of the site graph, consisting of only the pages visited and the links followed – the *navigation graph*. This navigation graph can contain information about one user session, about one particular user or about entire user groups. Although temporal information of the navigation path is not represented, many conclusions can be drawn from this model [12].

Attributes can be assigned to the pages and links within both the site graph and the navigation graph [1]. Attributes that reflect page content and user navigation can be used for assigning weights to links. These weights express relations between pages. Several weighting mechanisms for the site graph are mentioned in the literature:

- *uniform weighting*: each link receives equal weighting [7]
- *content-based weighting*: each link receives a weight corresponding to the textual similarity between the source and destination page (such similarity measures can be derived from the field of information retrieval) [7]

- *transition-frequency weighting*: each link receives a weight which expresses the probability a user will follow this link, based on statistics derived from an individual user's path [12], from multiple sessions [18] or from the whole user population

The same weighting mechanisms can be applied to the navigation graph. Note that the weighted navigation graph can be seen as a weighted site graph with minimum weight assigned to each link that is not followed during the observed period. A fourth mechanism, specifically applicable to the navigation graph, expresses user interest in the requested pages [15] and is believed to improve accuracy in predicting a user's path [7]

- *view-time weighting*: each link followed is assigned the mean time spent on the source page before following this particular link [15]

Whichever weighting method one chooses, several metrics can be applied for analyzing the graphs' structure and features. An overview of such methods will be given in the next paragraph. Naturally, the resulting metrics need to be interpreted to become useful. In this process, the intuitive meaning of the chosen weighting method should be taken into account. Interpretation of the metrics is also needed for distilling the underlying cognitive mechanisms for the observed behavior [9]. These mechanisms are needed to relate observed metrics to specific adaptation strategies, a subject which is researched in a separate track of our project PALS Anywhere[1].

Analyzing Graph Structures and Features

The site graph and the navigation graph reveal a wealth of information. This paragraph presents metrics and methods that can be used for deriving site characteristics, different functional types of pages and user navigation characteristics.

Site Characteristics

Below several metrics are presented that reflect structural characteristics of a web site. These characteristics reflect a web site's size, complexity, linearity, distance and clustering respectively.

The *size* of a web site – the number of pages and links – is one of the most obvious measures. The larger or more complex a site, the more structuring is needed to prevent disorientation [2]. The size may depend on the type of web site – e.g. a personal site, an institutional web site, a newspaper or a portal site, all of which serve different user goals [16].

The c*omplexity* of a site indicates the amount of freedom in navigation offered to the user. Complexity is most easily expressed as the ratio between the number of links and the number of pages [14]. However, theoretically larger sites can contain more links. Therefore it makes sense to take the maximum number of connections – each page links to all other pages in the site – into account. In [14] this measure is defined as the *net density*, which is the ratio between the actual number of links and the theoretical maximum. The distribution of links – or the total distance between all pages, compared to a theoretical maximum and minimum – is expressed by a *compactness* measure, as proposed in [2].

The *linearity* of a web site, expressed by the *stratum* metric [2], indicates the order for reading imposed by the author. According to Botafogo [2], linearity reduces complexity. A high stratum indicates that only few navigation paths are allowed, whereas a low stratum indicates that all pages are reachable from anywhere. Correspondingly, the number of cycles within a site is also a measure of linearity [14]. Since, intuitively, the size of the cycles makes a difference, the graph-theoretic notion of *circumference* (a graph's longest cycle) [5] might be useful.

Distance, the shortest path between pages, can be viewed as an indicator of navigation effort. According to Smyth [18], cutting down the distance increases user satisfaction dramatically. An important distance measure is a node's *depth* [2], its distance to the root – a site's home page. Deep pages might indicate low relevance pieces of information. Other distance measures are the site's *diameter* – the maximal shortest path between any two pages – and *radius*

---

[1] PALS stands for Personal Assistant for online Services; the project is supported by the Dutch Innovative Research Program IOP-MMI. Our research partners are TNO Human Factors and the University of Utrecht.

– the minimal shortest path [5]. Some researchers have proposed studying the average length of all shortest paths instead, the *average connected distance* [3].

A web site usually consists of different *clusters*, richly interconnected parts of the site graph dedicated to a topic with but a few links to other topics [13]. This is similar to the graph-theoretic notion *block*, a maximal non-separable sub graph, connected to the remaining part via one node, the *cutnode*. [5]. By releasing the constraints so that cross-references do not hinder the recognition of a cutnode, this feature can be put into use. Related to the notion of blocks is the *median* [5], the collection of nodes with shortest distances to and from other nodes. Using the median measure, strategic positions within a web site or within a block can be recognized. Conversely, the *periphery* [5] indicates all pages that form a dead-end.

Page Functions

Within most web sites, different types of pages can be distinguished based on the function they fulfill. Although most pages that merely facilitate navigation are usually marked with proper names (e.g. menu, glossary), it might be useful to identify them by their local structural properties – for structure reflects semantic relationships between web pages [2]. In [13] and [2] the following page types are discriminated by such characteristics:

- *home pages* are the first pages of a set of pages one would visit. The distance from the home page to other pages within the set is small. According to Botafogo [2] the number of links on the home page (its out-degree) should not be too large. Its content similarity between itself and the reachable pages would be high. Cutnodes, mentioned in the previous paragraph, form the home page of a part of the web site
- *index pages* point users to a large number of other pages. Its distance to these pages should be small and moreover – its out-degree (the number of links) should be large [2]. In [13] index pages that also serve as a home page, are called *source index pages*
- *reference pages* are typically used to explain concepts (a glossary), or to contain references or acknowledgments. Reference pages can be looked at as the converse of index pages – the number of pages that link to the page (its in-degree) is expected to be large
- *content pages* do not facilitate navigation, but deliver information or contain services. A content page is expected to have a small in- and out-degree. Also its size is expected to be somewhat larger than the pages mentioned above

Besides this distinction between page types, one could also try and find a measure for a page's importance. Pages can be important because of their content, or because they point to many important pages. In [11] these pages, termed *authorities* and *hubs* respectively, are distinguished using an algorithm that ranks both categories at the same time.

User Navigation

As mentioned before, one can model user navigation as an overlay of the site graph. According to Shahabi [15] this is the most detailed information that can be gathered in an automatic way. As browsing is an open-ended task, metrics based on the navigation path rather than the classic HCI measures such as task performance efficiency are needed [17]. Smith produces such measures for lostness in hyperspace, efficiency of navigation and user confidence in their own navigation abilities [17]. According to Smith, lostness is indicated by degradation of user performance, which can be observed by an increased number of revisited pages. Interestingly enough, in other research [12] it has been shown that more successful users made extensive use of a web site's structure by returning to home pages and index pages very often. The latter results are in line with cognitive theories, which claim that navigation improves as the mental model of a document develops [19]. These temporal features of user navigation strategy can be expressed by adapted *compactness* and *stratum* metrics, applied to a *transition-frequency* weighted navigation graph [12].

Apparently, subtleties such as page functions are playing a role as well. In [7] and [15] the importance of the time spent viewing a page is stressed. Since people display a variety of browsing behaviors – such as random surfing, task completion and information seeking [1] – the resulting navigation paths will show different characteristics. Moreover, as indicated before, the structure of these paths will be heavily influenced by the design of the web site.

Adaptation of Site Structure to the Mobile Context

The previous section described how metrics derived from site structure, page functions and user navigation paths can be used for identifying user needs for support in navigation and orientation in a wide range of user contexts. We hypothesize that these metrics can also be deployed for actually adapting navigation support to various mobile and desktop settings. On page level, these adaptations can be created using common adaptive hypermedia techniques, such as link ordering, link hiding, link highlighting, link annotation and direct guidance [4]. On site level, these adaptations can be seen as a personalized subgraph imposed on the site graph. This subgraph can be evaluated using the same metrics as described in this paper.

In order to increase user satisfaction many adaptive systems strive to reduce dialogue complexity [20]. This goal may be reached by removing cyclical links – therewith increasing a site's linearity – or redundant links – therewith decreasing a site's density. More explicit linkage to navigational landmarks, such as home pages, index pages and clusters, might prevent a user from getting lost in hyperspace. A reduction of the average connected distance, for example by explicit linking to central nodes – which need not necessarily be landmarks – might lessen user navigation effort.

Since user navigation strategy changes with the context of use, one needs more than a few metrics and adaptation strategies. Based on the above literature and supported by both cognitive and graph-theoretic notions, we are constructing a toolbox of structure-based metrics and adaptations which can be used – in combination with content-based approaches – to make navigation easier on a wide variety of devices, in both mobile and desktop settings. Although information providers are already adapting their content to mobile context, we firmly believe that the proposed adaptations on the client side will still be essential to bridge the gap between user needs in different situations and the information that is provided.

References

1. Anderson, C.R., Domingos, P., Weld, D.S.: Personalizing Web Sites for Mobile Users. *Proc. Of the 10th Intl. WWW Conference*. ACM, 2001. pp. 565-575
2. Botafogo, R.A., Rivlin, E., Shneiderman, B.: Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. *ACM Transactions on Information Systems, Vol. 10, No. 2.* ACM, 1992. pp. 142-180
3. Broder, A., Kumar, R., e.a.: Graph structure in the web. *Proc. Of the 9th Intl. WWW Conference*. ACM, 2000. pp. 309-320
4. Brusilovsky, P.: Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction*. Vol 6, No. 2-3. pp 87-129
5. Buckley, F., Harary, F.: Distance in Graphs. Addison-Wesley Publ. Comp., 1990. ISBN 0-201-09591-2
6. Buyukkokten, O., Garcia-Molina, e.a.: Power Browser: Efficient Web Browsing for PDAs. *CHI 2000 Conference Proceedings*. ACM, 2000. pp. 430-437
7. Heer, J., Chi, E.H.: Separating the Swarm: Categorization Methods for User Sessions on the Web. *CHI 2002 Conference Proceedings*. ACM, 2002. pp. 243-250
8. Herder, E., Van Dijk, E.M.A.G.: Personalized Adaptation to Device Characteristics. In: De Bra, P., Brusilovsky, P., Ricardo Conejo, R. (eds): *Adaptive Hypermedia and Adaptive Web-Based Systems*. Springer Verlag, 2002. pp. 598-602
9. Juvina, I., e.a.: Analysis of Web Browsing Behavior – A great potential for psychological research. In: Pribeanu, C., Vanderdonckt, J. (eds.): *Task Models and Diagrams for User Interface Design*. INFOREC Publishing House, Bucharest - Romania, 2002
10. Jones, M., Marsden, G., e.a.: Improving Web Interaction on Small Displays. *Computer Networks No. 31*. Elsevier Science, 1999. pp. 1129-1137
11. Kleinberg, J. M.: Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM, Vol 46, No. 5*. ACM, 1999. pp. 604-632
12. McEneaney, J. E.: Graphic and numerical methods to assess navigation in hypertext. *Intl. Journal of Human-Computer Studies 55.* Academic Press, 2001. pp. 761-786

13. Pirolli, P., Pitkow, J., Rap, R.: Silk from a Sow's Ear: Extracting Usable Structures from the Web. *CHI 1996 Conference Proceedings.* ACM Press, 1996
14. Rauterberg, M.: A Method of Quantitative Measurement of Cognitive Complexity. In: Van der Veer, G., Tauber, M, e.a.: *Human-Computer Interaction: Tasks and Organisation*. CUD-Publ., Roma, 1992. pp. 295-307
15. Shahabi, C., Zarkesh, A., e.a.: Knowledge Discovery from Users Web-page Navigation. *Proc. of 7th Int. Workshop on Research Issues in Data Eng. on High Performance Database Management for Large-Scale Applications*, IEEE Comp. Soc. Press, Birmingham, UK, 1997.
16. Shneiderman, B.: Designing Information-Abundant Websites: Issues and Recommendations. *International Journal of Human-Computer Studies 47 (1)*. Academic Press, 1997
17. Smith, P.A.: Towards a practical measure of hypertext usability. *Interacting with Computers Vol. 8, No. 4*. Elsevier Science, 1996. pp. 365-381
18. Smyth, B., Cotter, P.: The Plight of the Navigator: Solving the Navigation Problem for Wireless Portals. In: De Bra, P., Brusilovsky, P., Ricardo Conejo, R. (eds): *Adaptive Hypermedia and Adaptive Web-Based Systems*. Springer Verlag, 2002. pp. 328-337
19. Thüring, M., Hanneman, J., Haake, J. M.: Hypermedia and Cognition: Designing for Comprehension. *Communications of the ACM Vol. 38, No. 8.* ACM, 1995. pp. 57-66
20. Weibelzahl, S., Weber, G.: Evaluation adaptiver Systeme und Verhaltenskomplexität. *Proceedings of the 7th Workshop Adaptivität und Benutzermodellierung in Interaktiven Softwaresystemen, ABIS-2000*, 2000. pp. 105 - 115