Using Log Data to Detect Weak Hyperlink Descriptions

Vera Hollink, Maarten van Someren, and Bob J. Wielinga

Faculty of Science, University of Amsterdam
Kruislaan 419, 1098 VA Amsterdam, The Netherlands
{vhollink,maarten,wielinga}@science.uva.nl

Abstract. Users who visit a web page for the first time select links on the basis of the link anchors and the descriptions of the links that are provided on the page. If these descriptions do not give an accurate impression of the underlying pages, users can not make the right choice and select links that do not lead to their target information. In this paper we present a novel algorithm to automatically detect links with weak descriptions on the basis of usage information stored in the sites' log files. The algorithm distinguishes several types of weak descriptions and provides recommendations for how descriptions of each type can be improved.

1 Introduction

Users who browse unfamiliar web sites in search of information choose links on the basis of the link anchors and the text surrounding the links. This only succeeds if these descriptions accurately match the contents of the pages behind the links. If the descriptions are misleading, users often click links that do not lead to target information, which results in unnecessary long navigation times. High quality link descriptions are especially important for users of mobile devices for whom navigation is already slower. Moreover, on these devices choosing a correct link is more difficult as the small screens display less context for the links.

Finding link descriptions that are clear to all users and in all contexts can be an extremely difficult task. Link descriptions are usually designed with a particular user population in mind, but it often happens that other user groups visit the site as well. For these user groups the descriptions can be confusing. Moreover, the perspective of a visitor of a site is in part determined by the context in which the visit is made. For instance, users who visit a site in a work context have different information needs than users who visit the site from home. To create link descriptions that are optimally tailored to a users' needs, properties of the context need to be taken into account.

In this work we present a novel algorithm to automatically identify misleading link descriptions on the basis of the usage of the pages. Links with accurate and unambiguous descriptions are characterized by a usage pattern in which exactly those users who are looking for the pages behind the link, follow the link. Other usage patterns indicate that to some users the descriptions are unclear. The algorithm evaluates the usage patterns of links and classifies their descriptions as strong (clear) or weak (misleading). For weak links it determines the cause of the problem and explains in what way the descriptions need to be changed.

The methods in this work focus primarily on hierarchical link structures. These structures include, for instance, hierarchical menus, web directories and the WAP structures that are considered in [1]. For these link structures it is difficult to find good descriptions, as the descriptions of the intermediate nodes in the hierarchy must not only cover the contents of single pages, but the total contents of a set of pages.

2 Related work

Several methods have been developed that analyze log files and link structures and on the basis of the analysis recommend to add or remove certain links (e.g. [2, 3]). These methods evaluate the presence or absence of links, but not the link descriptions. Systems that add new links autonomously need to provide descriptions for the links. These systems usually use hand-made rules to find descriptions, such as using the page title or url (e.g. [4]).

Nakayama *et al.* [3] present an algorithm to detect page pairs that are similar in content, but that are not frequently visited in the same session. They suggest to add a link between these pages if the pages are not yet linked. If a link is already present, they conclude that the page layout must be adapted to improve the visibility of the link. They propose several methods to improve the layout, including changing the link anchor or the text preceding the link. A limitation of their work is that they can only detect weak links between pages that are very similar in content. The usage-based character of our method allows us to find weak links between pages that are related in terms of user relevance, but that have different contents.

Srikant and Yang [5] propose a method to discover the location in a web site where users expect to find certain target pages. They assume that users follow links to the location where they believe a target is located and backtrack when they find out that no link to the target is present at the expected location. Their method computes for each target page the positions where users frequently backtrack and recommends to add links to the target at these positions. This approach is similar to our approach in that both methods aim to determine incorrect navigation paths. However, Srikant and Yang search for the end points of these paths (the backtrack points). They solve the problems at the end points by adding links to the target pages. In contrast, we determine the source of the problem, the point where users deviate from the optimal path. The problem is solved at these points by improving the link descriptions that gave users incorrect expectations about the contents of the underlying pages.

3 Method

In this section we formalize the notion of a weak link description and present the algorithm to discover weak descriptions. In addition, we distinguish two types of weak descriptions and discuss how weak descriptions can be improved.

3.1 Preprocessing

Usage patterns are extracted from the data collected in server logs. Before the patterns are extracted, the log data is preprocessed. The sessions of individual users are restored with the method described in [6]. All requests coming from the same IP address and the same browser are attributed to one user. When a user is inactive for more than 30 minutes, a new session is started. After the sessions are restored, the pages in the sessions are classified into auxiliary and target pages. A page is a target page for a user if it provides a (partial) answer to his information needs. Auxiliary pages do not contain information that is interesting for the user, but only facilitate browsing. We use the classification method described in [6], which is based on the time that a user spent reading a page. All pages with a reading time longer than a reference length are marked as targets. The other pages are auxiliary pages. As reference length we use the median reading time of the terminal nodes of the hierarchy.

3.2 Detection of weak link descriptions

The algorithm for detecting weak link descriptions is given in Fig. 1. Below we describe the algorithm in detail. The following definitions are used:

Menu A whole hierarchical menu structure.
Menu fragment One node in a menu together with its direct children.
Link description The anchor text of a link and the text surrounding the link that gives information about the contents of the link.
Link content All content (text, images, etc.) that is located under the link in the hierarchy.
(In)correct link A link whose content contains (n)one of the user's target pages.

To evaluate the link descriptions in a menu, the menu is first divided in menu fragments. The descriptions are evaluated fragment by fragment (Fig. 1 (i)). For each fragment we count how many times a user with a target under each child link in the fragment opened each other child link in the fragment. These data are stored in a matrix. We call this matrix

Algorithm 3.1: EVALUATE_MENU($Menu$)

```
evaluations ← ∅
for each fragment in Menu                                              (i)
  do { Create confusion matrix C                                       (ii)
       for each row in C with correct link l
         do Add EVALUATE_LINK(row, l) to evaluations
return (evaluations)

procedure EVALUATE_LINK(row, correct_link)
 if BINOMIAL_TEST(correct_link, row) = too low                         (iii)
   then { confused ← ∅
          for each incorrect_link k in row
            do { if BINOMIAL_TEST(k, incorrect_links) = too high       (iv)
                   then Add k to confused
          if confused ≠ ∅
            then return (evaluation(correct_link, confused_with(confused)))
            else return (evaluation(correct_link, unclear))
   else return (evaluation(correct_link, strong))
```

Fig. 1. The link description evaluation algorithm in pseudocode.

a confusion matrix as it shows how often users clicked correct links and how often they confused links with other links (Fig. 1 (ii)). Fig. 2 shows an example of a menu fragment with four child items and a confusion matrix for this fragment.

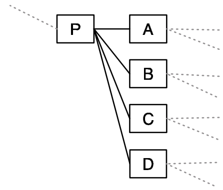
The link descriptions in a menu fragment are evaluated using the corresponding row in the confusion matrix. If a description is accurate, the frequency of the clicks on the correct link is high and the other frequencies are low. In other words, people are able to select the link that leads to a target. Large numbers of clicks on incorrect links indicate a weak link description.

We distinguish two categories of weak descriptions: unclear descriptions and confused descriptions. The first category includes descriptions that are unclear in itself. These descriptions do not match the contents of the link, so that users can not accurately predict whether their target information is located under the link. As a result, many users first try one or more incorrect links before selecting the correct link. In this case, in the matrix not only the frequency of the correct link but all frequencies in the same row are high. An example of a link with an unclear description is item D in Fig. 2: in the fourth row of the confusion matrix all frequencies are relatively high.

The second category of weak links includes links whose contents are in part covered by the description of another link in the same fragment. Users do not know which of the two links they should open and often open the incorrect link. In the matrix both the frequency of the correct link and the frequency of the other link are high, while all other frequencies in the row are low. In the example in Fig. 2, link B is often confused with A.

To determine whether a frequency is too low or too high, we use a statistical test. For each row in the confusion matrix we compare the number of clicks on the correct link to the number of clicks on incorrect links by means of a binomial test (Fig. 1 (iii)). If the proportion of clicks on the correct link is with 99% chance lower than the expected probability π, the description of the link is marked as weak. By default, the value of π is set at the median of the observed probabilities of all correct links in a menu. The value of π is always greater than 0, because not all deviations from the optimal path are the result of weak descriptions. In some cases, users may have abandoned their navigation paths, because they changed their information needs during browsing. The category of a weak link is determined by comparing the frequencies of the clicks on the incorrect links (iv). If the binomial tests show that an incorrect link has a significantly higher frequency than average, the correct link is confused with the incorrect link. If none of the incorrect links have a significantly high frequency, the description of the correct link is unclear.

We illustrate the link description evaluation procedure using the example in Fig. 2. First, the description of link A is evaluated. The first row of the confusion matrix shows that in total there were 109 (100+6+2+1) occasions in which users with a target under A made a



		Clicked link			
		A	B	C	D
	A	**100**	6	2	1
Target	B	40	**86**	0	1
under	C	0	6	**90**	2
	D	9	10	12	**61**

Fig. 2. Example menu fragment with four child items (left) and a confusion matrix with example frequencies (right). Clicks on correct links are shown in bold.

navigation step within the menu fragment. 100 of these users selected link A. The binomial test shows that this is not significantly lower than expected. Therefore, we conclude that A has a strong description. In 86 of the 127 occasions where the target was under B, B was chosen. The test indicates that this is too low and we conclude that the description of B is weak. Next, we compare the frequencies of the clicks on incorrect links: 40, 0 and 1. The value of 40 of link A appears to be significantly higher that the other two values, which indicates that surprisingly many people with a target under B click link A. Therefore, we say that link B is confused with A. Testing the third row indicates no problems with the description of link C. Link D is chosen on only 61 out of 92 occasions, which is significantly low. If we compare the frequencies of the clicks on the incorrect links, we find that none of them is too high. Consequently, the description of link D is classified as unclear.

Until now we explained how link descriptions can be evaluated for a user population as a whole. As stated in the introduction, in some domains it is important to create different descriptions for different groups of users or for users in different contexts. This can be accomplished by first clustering the log data in a number of user clusters with similar navigation patterns or similar contextual properties. Several methods have been developed for this purpose (e.g. [7]). Once the data is clustered, a link description analysis is performed for each cluster using only the sessions of the users from the cluster. In this way, it is possible to find link descriptions that are clear to one group of users but unclear to others.

3.3 Improving weak link descriptions

Once we determined which anchors are insufficient, a web master needs to improve the problematic links. The system provides several guidelines to solve the various problems.

A description can be unclear because some of the contents of the link do not fall in the category that is suggested by the description. This problem can be solved by giving a broader description to the link. Another solution is to maintain the description and group the items that are not covered under a new menu item. A second cause of unclear links is the use of terms in the description that are not known to the users or that have a vague meaning. In this case the description needs to be reformulated in terms used by the user community. One source of such terms are the query terms submitted to a search engine as these terms are typed in by the users themselves.

If a link A is often confused with link B, part of the contents of A are covered by the description of B. This can be solved by assigning a narrower description to B. Alternatively the content items under A that are most frequently confused can be moved or copied to B. If the confusion can not be attributed to some specific content items, A and B can be merged into one large item.

4 Evaluation plan and preliminary results

The link evaluation algorithm was applied to the log files and menus of three Dutch web sites. The menus consisted of 85 to 287 links, 9-21% of which were marked by the algorithm as having weak descriptions. On all sites both unclear and confused descriptions were found.

Inspection of the link descriptions that were classified as 'unclear' showed three main causes for weak descriptions. Some descriptions appeared to be too general. For instance, on a site for elders about the prevention of falling accidents one of the menu items is called 'hints for a safer home'. This description is too general as it covers almost the entire topic of the site. Another cause of weak links is the use of jargon words that are not commonly understood, such as 'grab pole' and 'threshold ramp'. Finally, few descriptions consisted of ambiguous terms where the less common meaning was intended. Links that were confused with other links often had closely related descriptions, such as 'care providers' and 'care institutions'. As these items are very similar, a good solution would be to merge them.

These results are promising as they show that the algorithm is able to identify a number of important shortcomings of link descriptions. However, they do not show the effectiveness of the suggested improvements. To evaluate the practical value of the method, we are planning to conduct a user study. The link evaluation algorithm will be applied to the menu of a web site. On the basis of the suggested improvements the web master of the site adapts the link descriptions and/or the structure of the menu. A number of participants will be asked to search for certain information via the site's menu. Half of the participants use the original menu and the other half use the improved version. During the search assignments we measure the number of times an incorrect link is followed. Comparison of the results of the two menus allows us to measure the effects of the link evaluation algorithm in a realistic setting.

5 Discussion

From usage data we can compute how descriptive links are in their current context. However, we can not evaluate link descriptions per se as descriptiveness is highly context dependent. For instance, a description 'seal' is perfectly descriptive among 'stamp' and 'envelope', but when a link named 'walrus' is added it becomes unclear. Consequently, when changes are made to the contents or structure of a menu, a new link description analysis needs to be performed for the modified branches.

We are currently extending the classification scheme for weak link descriptions. The extentions allow us to recognize more subtile problems and to give more focused recommendations for improvements. Later, the link evaluation algorithm can be combined with algorithms to estimate navigation time, so that we can predict the effects of modifying the link descriptions and the menu structure on navigation time. These estimations can help a site master to make a choice between various possible solutions for a weak description. Another direction of future research are methods to find alternative link descriptions automatically, for example, by borrowing techniques from keyword extraction or text summarization. In addition, content analysis methods may be used to analyze why descriptions are weak.

The current work addresses only virtual navigation patterns in web contexts, but the presented algorithm can also be applied to physical navigation patterns. If we can track the movements of visitors in, for instance, a physical store, we can evaluate their walking patterns and determine the points where people frequently take wrong turns. This allows us to make recommendations for the improvement of the organization of the store.

References

1. Smyth, B., Cotter, P.: Intelligent navigation for mobile internet portals. In: IJCAI'03 Workshop on AI Moves to IA: Workshop on Artificial Intelligence, Information Access, and Mobile Computing, Acapulco, Mexico (2003)
2. Wang, Y., Wang, D., Ip, W.: Optimal design of link structure for e-supermarket website. IEEE Transactions: Systems, Man and Cybernetics - Part A **36**(2) (2006) 338–355
3. Nakayama, T., Kato, H., Yamane, Y.: Discovering the gap between web site designers' expectations and users' behavior. Computer Networks **33**(1–6) (200) 811–822
4. Schilit, B., Trevor, J., Hilbert, D.M., Koh, T.K.: Web interaction using very small internet devices. Computer **35**(10) (2002) 37–45
5. Srikant, R., Yang, Y.: Mining web logs to improve website organization. In: Proceedings of the 10th International Conference on World Wide Web. (2001) 430–437
6. Cooley, R., Mobasher, B., Srivastava, J.: Data preparation for mining world wide web browsing patterns. Journal of Knowledge and Information Systems **1**(1) (1999) 5–32
7. Mobasher, B., Dai, H., Luo, T., Nakagawa, M.: Discovery and evaluation of aggregate usage profiles for web personalization. Data Mining and Knowledge Discovery **6** (2002) 61–82