Explaining difficulty navigating a website using page view data

Paul Thomas
CSIRO
paul.thomas@csiro.au

ABSTRACT

A user's behaviour on a web site can tell us something about that user's experience. In particular, we believe there are simple signals—including circling back to previous pages, and swapping out to a search engine—that indicate difficulty navigating a site.

Simple page view patterns from web server logs correlate with these signals and may explain them. Extracting these patterns can help web authors understand where, and why, their sites are confusing or hard to navigate.

We illustrate these ideas with data from almost a million sessions on a government website. In this case a small number of page view patterns are present in almost a third of difficult sessions, suggesting possible improvements to website language or design. We also introduce a tool for web authors, which makes this analysis available in the context of the site itself.

Categories and Subject Descriptors: H.5.4 [**Information Interfaces and Presentation**]: Hypertext and Hypermedia
General Terms: Human Factors; Measurement
Keywords: Web documents

1. INTRODUCTION

For a large number of organisations, clear communication on the web is important: there is an imperative to help visitors find the information they want, lest they go elsewhere (e.g. to a competitor) or use some other, more expensive, channel (e.g. telephoning a helpdesk). Clearly, it can be a great help to understand site performance, visitor characteristics, and visitors' movements and experiences online.

As usual, if we want to understand this we have several options available. Involving users directly by observing them in action, administering surveys and questionnaires, or running interviews has the advantage of rich results accompanied by explanations of users' thought processes and responses; however, these techniques demand significant time and money and it is hard to get quick updates following website changes. The common alternative is to use records already available to understand a site: this is the goal of web analytics.

Typical web analytics is based on a web server's transaction logs. These are sparse, and contain only trace data, but they are readily available. From this, analytics packages[1] can provide an abundance of detail about a site, such as the number of visitors and where they are from; the number of page impressions; types of browsers; or the number of people who perform some action, such as making a purchase or clicking an advertisement (see e.g. Jansen [4] for an overview). These measures represent the mechanics of the web, but do not directly speak to user satisfaction, confusion, or other experiences.

In this work we attempt to bridge the two worlds, with analysis that is cheap, fast, and based on real usage; but that also provides some insight past a tally of pages visited.

We believe that server logs—aggregated appropriately—can provide insight into users' experiences online as well as the more mechanical view. We are interested in understanding what it is that this recorded behaviour can tell us about user experience, and how it can help web authors identify and fix problem spots in their sites. By contrast to conventional web analytics, we are trying to (1) *find common behaviour* that may (2) *explain why users are struggling* to find information; then (3) *present that* information to web authors in a way they find useful.

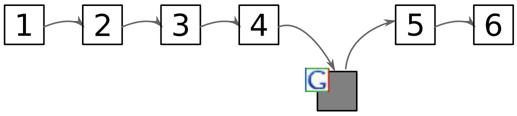
2. OUR APPROACH

We approach this problem through analysing user sessions as recorded in web server logs. Logs provide very simplistic data—they record only page loads, not interaction with the page, and cannot make important distinctions such as that between the time a user spends reading and the time they spend away from the screen. Despite this, they are easy to obtain for any website, and they can be collected without any extra tools on users' (or publishers') computers.

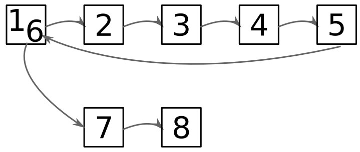
Rather than individual page views, we consider complete sequences of views—"sessions". This is for two reasons: first, there is very little information in a single recorded page view. Second, sessions describe the whole of a user's experience on a site—or as close as we can get via standard logs—which for us is more important overall than their interactions with a single page.

[1] Popular packages include those from Google (`www.google.com/analytics`), Yahoo! (`web.analytics.yahoo.com`), Clicky (`getclicky.com`), and WebTrends (`webtrends.com`). Open-source offerings include Piwik (`piwik.org`).

ADCS '12, December 5-6, 2012, Dunedin, New Zealand.



(a) A "swapping" session is one that includes one or more page views on our site (page views 1–4 in this example), a visit to a search engine (between 4 and 5), then one or more views on our site (5 and 6). The first portion (1–4) we call the "pre-swap" phase.



(b) A "circling" session is one that includes a sequence of page views (here 1 to 5), then a return to a previous page (6). The first portion (1–5) we call the "pre-circle" phase.

Figure 1: Two indicators that a user is having difficulty navigating a website: "swapping" and "circling" sessions.

For example, a user may visit some page p but only after issuing a search and visiting several other pages. That search, and the other pages visited, suggest something about what the user was trying to do with p and what task they had in mind; if we only look at transactions immediately preceding p, rather than the whole session, we would miss these clues. Similarly, the final page in that session may be the one that satisfied the user's information need. Knowing where that was, and not just which page immediately followed p, might provide clues for site design.

Our approach has three parts. First, we label those sessions where a user is having trouble: to do this, we need to identify some signals of navigation difficulty that we can extract from web server logs (Section 3). Second, we examine the logs to find patterns of behaviour that are relatively common in these sessions, ideally ones that happen as early as possible: that is, we need to find patterns that help to explain why some users have trouble (Section 4). Finally, we expose this to web authors to help them refine their sites (Section 5).

Section 6 discusses other approaches and related work.

3. LABELLING SESSIONS

We base our session labelling on web logs from a large Australian government agency. These logs include use of both the agency's intranet and external sites, and individual users can be distinguished [12].

In around 5% of cases, agency staff started on one of the two sites, spent some time navigating around, then switched to the other. From discussions with staff, discussions with web authors, and from examining these sessions we believe that in a large number of these cases staff were struggling to find information on the first site—probably because they were looking in the wrong place.

The first parts of these sessions must have involved some level of frustration. Comparing behaviours in these parts with behaviours in simpler sessions did show up some differences. Two behaviours in particular, "swapping" and "circling", were much more common when a user was about to change sites. The first is defined as follows:

> A *swapping* session is one in which a user browses our website, exits to a search engine, performs a search, then returns to our site (Figure 1a).

We believe this is clear evidence of navigational trouble: rather than follow links on the site, the user has found it easier to use an entirely separate tool.

Note that if the first page view in a session is the only one referred from a search engine, that session is not "swapping" by definition. Also, since we use the HTTP `Referrer:` header to recognise swaps, we are unable to detect cases where a user swaps to a search engine but never returns, for example because their information need was satisfied by the engine's snippet. Nor are we able to recognise swaps where the HTTP header is absent (about 13% of cases in the data we have examined). Our count of swapping sessions is therefore in some regard an under-estimate.

We call the part of a session before the swap the "pre-swap" phase.

Our second signal is "circling":

> A *circling* session (Figure 1b) is one in which a user views a number of pages, then retraces their steps and re-views an earlier page.

This may indicate, for example, that a particular link did not lead to the expected information; or that information the user needs is split between several pages with no obvious path between them.

Any number of repeated visits, to any page (not just the first), marks a circling session. The part of a session before the first re-view we call the "pre-circle" phase.

There are some circumstances where circling might be expected—for example, users may return again and again to a list of open jobs—so this is a weaker signal than swapping and must be interpreted with some knowledge of the site and its users. The signal is still useful, however, as the illustrations in Section 4 show.

The two behaviours are therefore suggestive, if not perfectly diagnostic, of difficulty navigating a website. In the present work we use these as indicators of "struggling" sessions, which are sessions where a user is having trouble finding the information they need. We label as struggling those sessions that are swapping, circling, or both.

Further signals, based on the data in server logs, are certainly possible (for example, one might consider reloading a page to mean something). One can also easily imagine signals from augmented logs—perhaps injecting JavaScript to log mouse movements or scrolling, for example [5]. In this work we are concentrating on logs available anywhere, from any server; crucially, this means our analysis does not depend on any modification to the site or to the server itself.

4. FINDING PATTERNS

If we believe that swapping is a sign of a user's navigational difficulty, it would be useful to know whether there are certain user behaviours, or certain pages, that tend to provoke it—and similarly for circling. For example, people who look at certain pages may be more likely to swap out to Google to find what they are looking for, rather than follow the authors' cues; or people who follow certain links may be more likely to retrace their steps to gather information or to try another path. In particular, we ask:

> Given logs from a web server, are there patterns of user behaviour in the pre-swap or pre-circle phases that predict swapping or circling sessions?

That is, are there behaviours that tend to occur before circling or swapping, and are relatively more likely in these sessions? For example, in Figure 1b, is there some feature of user behaviour in 1–5 (or 1–4, or 2–5, etc.) that tells us the user is likely to circle back and try again?[2]

If the answer is "yes", then we believe these patterns could be a good clue to what is confusing a user. Of course we do not attempt to replace expert analysis with an automated tool: that said, being able to extract accurate patterns should help web designers and authors identify troublesome parts of their website, particular pages, or even particular links, and edit or redesign them. In this section we present simple techniques to do this, and briefly describe the sorts of patterns we look for at present.

The techniques are general and we believe they can work on a range of websites and can provide useful pointers for authors and analysts. We illustrate them, however, with a case study using data from almost a million sessions on a large government website. Based on the presence of swapping or circling, sessions were partitioned into "good" and "struggling" sets; the objective then is to find differences between the two that suggest where and why users have trouble.

4.1 Illustrative data

The website is that of a large government agency (not that of Section 3) that administers a range of programmes and has dealings with around a third of the population. The site is aimed mostly at individual citizens and residents, but it includes sections for other audiences such as businesses and professionals; it also contains a large number of forms and booklets in PDF format. This site serves several audiences and is intended to replace face-to-face or telephone enquiries as much as possible, so simple navigation is important. Authors have put some effort into arranging pages so there is a single page to answer most expected needs, and so that page can be found quickly.

We obtained one week's worth of log files from the site, representing just under one million sessions. URLs were normalised, and records that appeared to be robots were removed (including records of sessions more than two hours long, or requesting more than 1000 pages).

Table 1 summarises the cleaned data. In this table, "pages in site" counts pages actually visited, after minimal normalisation; the server hosted more pages, but these were not visited during the week. A "user" is a combination of browser and IP address, and a "session" is any sequence of page views from the same user, with no more than 30 minutes between consecutive views. "Struggling" sessions, those that swapped or circled, were truncated to the pre-swap or pre-circle phase, which means in these cases the records are of users' behaviour as they are just starting to have difficulty. (From conversations with the relevant web authors and other staff, we are confident that the two signals discussed above are appropriate for this site as well.)

[2]Note that this is an off-line problem, and we are not trying to predict or detect a user's difficulty and intervene in realtime. Even if we could manage the extremely high precision needed to intervene only when it's actually needed, it's not clear what form that intervention should take. In this work we are only interested in analysis for web developers.

Log length	7 days
Pages in site	7,345
Users	748,099
Sessions	990,600
Good sessions	82%
Struggling sessions	18%
Mean views, good sessions	1.4 ± 1.3
Mean views, pre-swap/circle	2.2 ± 2.2
Mean duration, good sessions	$0{:}54 \pm 3{:}32$
Mean duration, pre-swap/circle	$2{:}08 \pm 5{:}15$

Table 1: Summary statistics. Views and durations are shown as mean ± one standard deviation.

4.2 Session characteristics

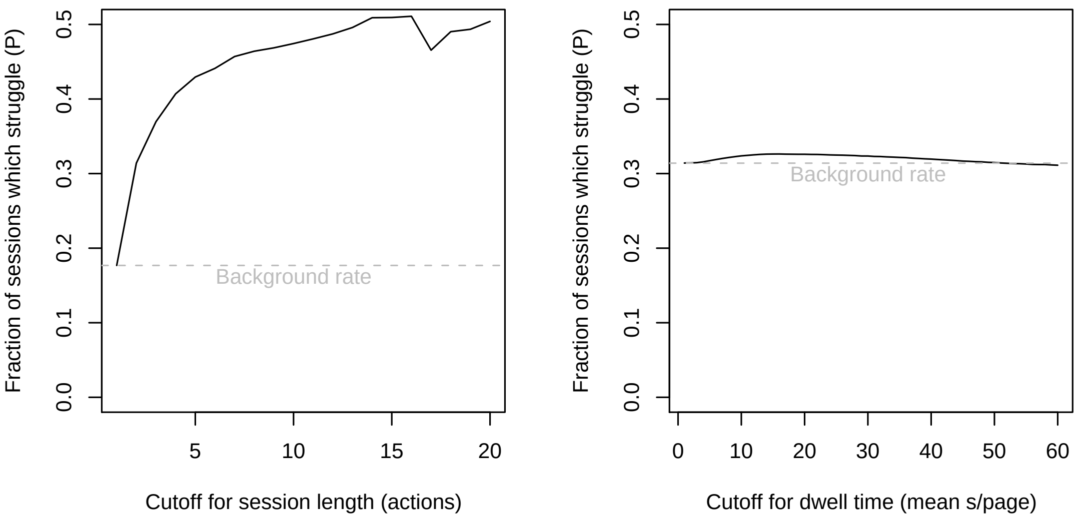
One trend is immediately apparent from Table 1. Struggling sessions are longer, in page views and time (despite being cut off just before the swap or circle), and the difference is significant (one-tailed Welch's t test, $p \ll 0.01$). They are also much more variable than good sessions. An obvious implication is that it could be useful to examine sessions with many page views, sessions that take a long time, or both.

Illustration. We illustrate the idea with our government website data. Figure 2a plots the fraction of sessions that go on to circle or swap (precision, vertical axis) against minimum session length (horizontal axis). There is clearly an effect: users who are viewing their third page in a session have a 37% chance of going on to swap or circle, more than double the base rate, and this rises to almost 50% at 13 page views before levelling off.

Unfortunately, this is of limited use. Most sessions on this site are short, so although a long session is more likely to lead to the behaviours we associate with navigation difficulty we are not able to explain many struggling sessions this way. (With a cutoff of five pages, recall is only 12%: that is, only one in eight struggling sessions will be captured.)

Elapsed time was not useful to distinguish good from struggling sessions, again on this data. However, since struggling users tend to view more pages (Table 1), we also considered classifying sessions based on the time spent on each page (Figure 2b). (Note that since at least two page views are needed to derive dwell time, this is restricted to sessions with two or more page views. The background rate is therefore about 31%, not 18%.)

As expected, sessions with lower mean dwell times are more likely to belong to users who are having trouble and are about to swap out to an external search engine or to retrace their steps. When dwell times approach one minute per page, struggling sessions are less likely. This is consistent with past observations on both search and browsing behaviour [3, 6, 10]. Gains over the baseline are not large, however: the



(a) Struggling sessions as a function of session length, in page views: the fraction of sessions which go on to swap or circle if they have x or more page views

(b) Struggling sessions as a function of dwell time: the fraction of sessions which go on to swap or circle if the mean dwell time is $\geq x$ s/page

Figure 2: The rate of occurrence of struggling sessions, as a function of session properties.

correlation between the chance of struggling and mean dwell time is poor.

4.3 Patterns of page views

A simple rule based on session length does detect struggling sessions early, but—at least in our example—it is not useful for authors or designers. It does not capture many struggling sessions (that is, recall is low). More importantly, it has little explanatory value: even when we know a user has had trouble, it is not clear where in the session they may have become confused. Instead, we consider slightly more complex patterns, which have the advantage of providing more explanation. In particular, we start by examining sequences of page views.

For each session in each of the "good" and "struggling" sets, we enumerate all sequences of page views—so for the three-page session $\langle x, y, z\rangle$ we would consider sequences $\langle x\rangle$, $\langle y\rangle$, $\langle z\rangle$, $\langle x, y\rangle$, $\langle y, z\rangle$, and $\langle x, y, z\rangle$. (In the examples here we ignored sequences of length five or more, and as always in the "struggling" set we only consider pages before the first swap or circle.)

Counting where these sequences appear allows us to calculate, for each sequence, the number of struggling sessions featuring the sequence; the number of good sessions; the precision P of this sequence (that is, the proportion of sessions with this sequence that do show signs of struggling); and the recall R of this sequence (the proportion of all struggling sessions that have this pattern). We can also calculate p for the hypothesis that the sequence was no more common in struggling than in good sessions (χ^2 test with Yates's continuity correction).

A sequence that is significantly more common in struggling than in good sessions, with reasonable precision and recall, may help explain why and where visitors have difficulty. We illustrate this below.

Illustration. There are 1573 sequences in our sample data where P > 0.5 and $p < 0.05$, some of which are shown in the top part of Table 2.

Some patterns are very precise: that is, a very high proportion of sessions exhibiting some patterns go on to signal difficulty. This is a strong clue that the pattern somehow explains users' confusion. Unsurprisingly, the most precise signals tend to be very particular. For example, in 77% of cases users with the pattern of row 1 (Table 2) went on to struggle, more than four times the expected rate; but the pattern was only seen in 22 sessions. Understanding this pattern and fixing the website accordingly would most likely be very helpful (high P), but would only fix a small number of sessions (low R).

In some cases a single page view correlated well with our signals of difficulty. The second example in Table 2, "claim form for card x"[3], is a case in point: 64% of users who looked at that page went on to circle or swap. This should suggest to an author that the form and surrounding information needs work. Quite likely the form itself is not self-contained, and further information is needed to fill it out (so users have to backtrack); or perhaps links to the form somehow attract users who don't in fact want it.[4] Note that conventional web analytics, based purely on counting page views or with very simple models of sessions, could not distinguish this troublesome page but could only report 1025 successful downloads.

Row 3 ("programme y overview $\rightarrow$ programme y eligibility") is typical of many of this type, with precision from 0.52 to 0.61

[3]These examples are anonymised.

[4]We could distinguish these two cases, and further explain users' difficulty, by looking at the page views immediately prior. If there is some page p such that the transition $p \rightarrow$ "claim form for card x" is relatively common in struggling sessions, then the link on page p is worth investigating. The software described in Section 5 supports this sort of ad-hoc analysis.

	Page view pattern	n_s	n_g	P	R
1.	income and asset tests → income tests → payments	17	5	0.77	< 0.01
2.	claim form for card x	652	373	0.64	< 0.01
3.	programme y overview → programme y eligibility	210	137	0.61	< 0.01
4.	forms starting with "f"	185	177	0.51	< 0.01
	Subsite view pattern	n_s	n_g	P	R
5.	paper publications	5103	9172	0.36	0.04
6.	audience z → paper publications	926	1283	0.42	0.01
7.	payments → factors affecting rates and eligibility	2461	3470	0.41	0.02
8.	payments → payments (different pages)	20147	30487	0.40	0.15

Table 2: Sample patterns which tend to occur in a struggling session. n_s and n_g are the number of occurences in the pre-struggle (s) or good (g) set; P is precision, i.e. the fraction of sessions with this pattern which go on to circle or swap; R is recall, the fraction of struggling sessions which contain this pattern.

for different programmes. This suggests a general problem. This pattern may appear if, for example, there's no link to apply for a programme from the eligibility page—having found they're eligible, people would have to backtrack to find out what to do next. Casual inspection of several "eligibility" pages suggests that is in fact the case, and suggests a simple fix for web authors to improve users' experience.

Finally, row 4 suggests that people have trouble navigating long lists. In fact there are 22 such lists of forms, for every letter except "g", "o", "x", and "z"; sessions passing through any one of these pages will struggle with probability well above normal. Again, this is a strong hint that some redesign is needed, although in this case the "better" design is not quite so immediately clear.

Other patterns were also apparent in our test data—for example, many common sequences involved different pages about asset testing or financial arrangements, information that would be clearer if presented on a single page. Several other patterns involved users switching between pages on two very different programmes that happen to have similar target audiences: this suggests users may be confusing the two and getting lost.

Note that the patterns we extract do correlate with struggling sessions to an interesting degree, and in many cases do suggest changes to the website. In general however the patterns could be *causes* or merely *symptoms* of confusion: for example, users viewing an FAQ page are likely already lost, and "fixing" the FAQ may not help. It may be possible to tell these two cases apart, if one pattern consistently occurs before or after others. At any rate, an analyst must still apply their knowledge of their website and users.

4.4 Patterns of subsites

The patterns of page views we extract with the method above can suggest particular navigation difficulties, but the same analysis can be run on different sequences. In particular, in many cases a web site will comprise several subsites, and if there are interesting patterns in subsite visits this might suggest problems with entire themes or topics as well as individual pages or links. With some definition of "subsite", which will vary for each web publishing or serving technology, the techniques above can be employed unchanged.

Illustration. Again we illustrate the idea with data from our government site. In this case, web authors took some care with URLs, and it was possible to define "subsite" simply as a page's containing directory (so a page `a/b/c.html` would be in the subsite `a/b`). Subsites defined this way were each specific to a topic, a particular segment of the audience, or a particular task.

In this data, there are indeed particular subsites that tend to lead to circling, or to people swapping out to a search engine. The bottom half of Table 2 contains examples, where P is significantly higher than the background rate ($p < 0.05$).

Again these patterns suggest areas where users are having difficulty. For example, 36% of people who visit the "publications" subsite (PDF forms of paper publications) went on to swap to a search engine or doubled back to find something else (line 5), rising to 42% if they came from a page specific to audience z (line 6). (Recall that the base rate is 18%.) Either they are unable to find the single correct form, or several forms are needed. In either case redesign might be called for. Visitors going from "payment information" to "factors affecting rates and eligibility" struggle 41% of the time (line 7). It seems likely that the directories of paper forms are confusing; and, further, that they are more confusing for audience z than for other audiences.

Unlike page-based patterns, some of these subsite-based patterns have high recall. Movements between any two pages in the "payments" subsite (line 8) occur in 15% of all struggling sessions before the point of swapping or circling, with 40% precision. Patterns such as this might suggest an entire section of the website is laid out in a way that doesn't correspond with users' expectations—for example, maybe information that belongs together is split between several pages, or the language used doesn't match users' expectations. Closer examination of pages in this subsite would prove fruitful.

4.5 Combinations and triage

Patterns of page views or subsites can be fairly precise—that is, they can correlate well with signals of struggling sessions and provide some indication why users are having difficulty—but they tend to capture few sessions. In our illustration the single best page view pattern has recall of only around 8% and most have recall much less than 1%. Individual subsite patterns capture more sessions, but in general recall is still low.

It makes sense for web authors and analysts to focus their attention on those pages, subsites, or patterns that explain as many struggling sessions as possible: if analysis

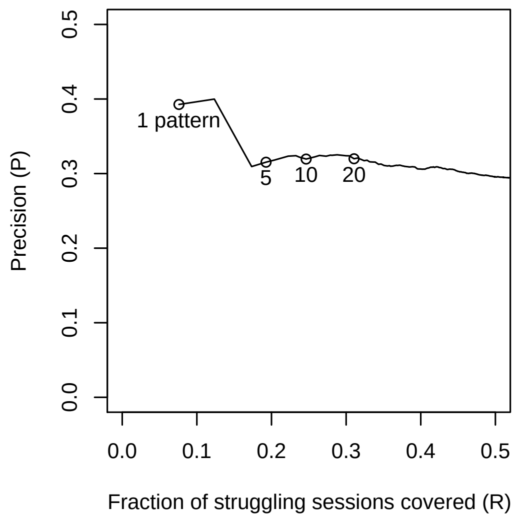


Figure 3: Precision and recall as more page view patterns are considered. Five patterns suffice to capture 19% of struggling sessions, and ten capture 25%, still with good precision.

suggests a fix, this will maximise the number of visitors who benefit. We do this by presenting *combinations* of patterns, to increase recall, and ordering them by *impact*, to help with this triage. In particular, we suggest a simple rule. First, take all those patterns that occur significantly more frequently before swapping or circling; second, prune the list to keep only those with reasonable precision; finally, order them from highest to lowest recall. Web authors should pay attention to the patterns at the head of this list.

Illustration. In our example, we set the threshold for "reasonable" precision to 0.2—that is, we discard patterns with precision < 0.2. The remaining patterns, ranked by recall, give the precision/recall curve in Figure 3.

Recall at the head of the list is good. Considering as few as five page view patterns covers almost a fifth of struggling sessions: clearly authors will not be able to fix all the difficulties in all these sessions, but any improvements will go a long way. With ten patterns, recall is 0.25; and with twenty patterns, recall is 0.31 with precision still 0.32. It seems reasonable to ask web authors, who know their site, to look at five to twenty patterns, and any improvements based on these patterns could improve many users' experiences.

5. EXPOSING INFORMATION FOR WEB AUTHORS

The ideas above—session-level web log analysis, finding signals of "struggling" sessions, and highlighting visitor patterns that are statistically more likely in such sessions—are instantiated in prototype software, presently deployed with the agency of Section 4.1. Again, the goal is not to automate anything: we still need people who understand their site and its users. Rather, we hope to present our analysis in a way that is useful for web authors charged with developing and maintaining a site.

As well as a backend that digests web server logs, the "latte" software includes several tools for web staff. The most important of these is the "web authors' sidebar", illustrated in Figure 4. The sidebar shows highlights of the analysis and pops up whenever a page on the relevant site is being viewed: the key idea is to integrate the analysis directly into the website, so analysts and authors can see the relevant numbers in context, while seeing what their users see. This contrasts with other tools, where the analysis is on a separate web page, spreadsheet, etc., and extra effort is needed to switch back and forth and make connections with what users see. Web authors have been using the sidebar as they move quickly about their website, looking closer when the data seems interesting.

The sidebar includes summary statistics of sessions through the page, and the proportion of those that are struggling according to the signals of Section 3. (Latte also uses two other signals, "long" and "slow" sessions, not described here.) If the fraction of struggling sessions is significantly higher than expected, either overall or by any one of those signals[5], it is marked with a "caution" sign, "**!**" (labelled (a) in Figure 3). This should suggest to the analyst that the page itself, or a pattern including this page, is causing visitors difficulty.

The sidebar also highlights other data that we think are worthy of attention. For instance, where we can glean search terms from HTTP `Referer` headers they give an explicit indication of what users were thinking of in a session. Any mismatch between the terms here and the terms on a page is likely to indicate difficulty, for example caused by visitors using different terminology or conflating two different ideas. Again, we mark with "**!**" those search terms which, when used in a session that included the present page, tend to occur when users are struggling (labelled "(b)").

As each page is viewed we also present the most common prior and following pages—that is, the most common patterns of length 2 that include the present page. Patterns that are significantly more likely to be in a struggling session are again marked (see label (c)). Showing the data in the context of the page itself makes these patterns more explicable: for example, instead of showing URLs or page titles we can show the text of appropriate links. In our experience this combination of flagging short patterns that correlate with navigational difficulty, and presenting them in context, often provides strong hints on awkward wording or confusing link structure.

We also include a tally of session endpoints (not shown in Figure 4), again for sessions passing through the current page, and mark those that are suspicious. The intuition is that the final page is often where users gather the information they need: presenting this to analysts can help clarify users' intentions in the same way as search terms.

A more detailed report is also available for each page, including session entry points, more patterns, and other data.

Finally, latte presents a list of "worst offenders". The "worst offenders" list provides a triage function similar to that described in Section 4.5: it ranks individual pages according to the number of struggling sessions through each page, provided the overall proportion is significantly higher than the background rate. Again, this helps web analysts focus on the fixes that will help the most people.

The prototype is in ongoing development, and we are working on ways to expose data on longer sequences, subsite-level

[5]To stop the signs being overwhelming, we have found it useful to raise the threshold for our hypothesis tests and only flag those cases where $p < 0.005$.

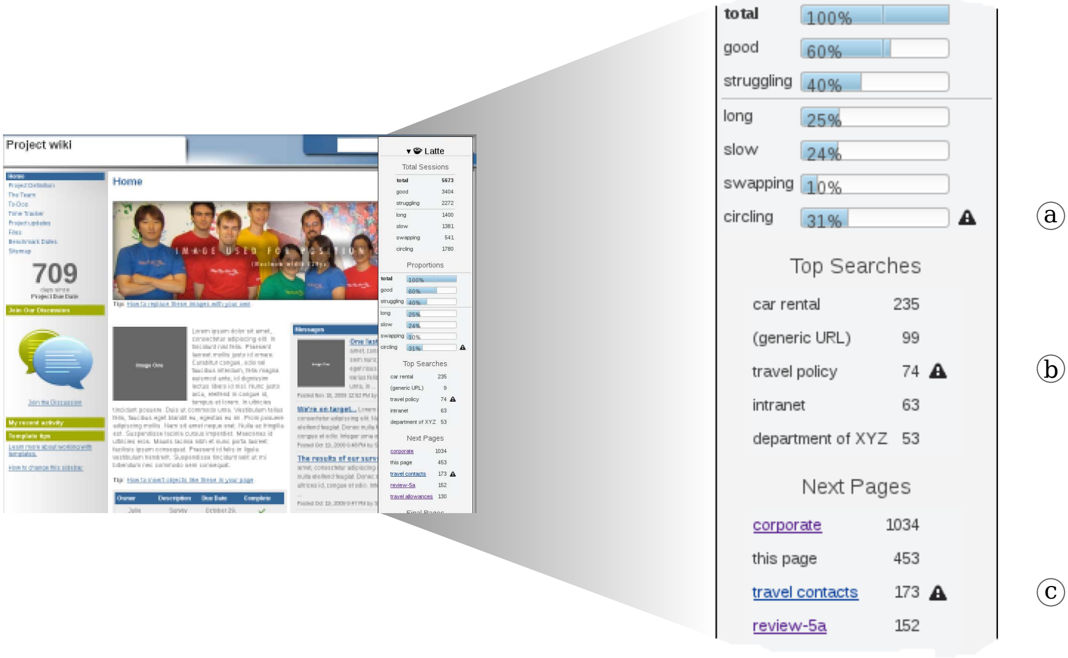


Figure 4: The "web authors' sidebar" in latte provides hints while an analyst is viewing a webpage. Here, (a) sessions that include this page are significantly more likely to circle than sessions on the whole; (b) users who search for "travel policy" in sessions through this page are significantly more likely to struggle; and (c) the sequence $\langle$ this page $\rightarrow$ travel contacts $\rangle$ is significantly more likely in a struggling session. (Page names and search terms are anonymised.)

analysis, and trends in over time. Nevertheless, on feedback to date the tool is useful, and staff at our collaborating agency have used it to prompt changes to both navigation and wording of their site.

6. RELATED WORK

Other research has suggested similar uses for web log analysis, and we briefly sample some related work here.

The approach we suggest is somewhat similar to Fox et al.'s "gene analysis" [3], where abstract codings of sequences were used to help predict success at the end of a session. In this work, by contrast, we are trying to explain difficulty and we report very particular sequences—down to an individual page or even an individual link.

Sequence mining algorithms learn common patterns from recorded sequences, such as server logs (a recent survey is that by Mabroukeh and Ezeife [7]). Spiliopoulou et al. [11] use such techniques to extract abstract sequences of page views by a website's "customers" and "non-customers", and compare the two sets for leads on how to increase conversions. This is a similar idea to ours, although we are interested in a different partition, but the focus is different: while Spiliopoulou et al. involve web analysts closely in pre-processing and pattern mining, we suggest a lighter-weight, more generic tool that is appropriate for information gathering as well as transactional tasks.

Other approaches include that of Nakayama et al. [8], who suggest that pages with similar text should be visited in the same session and that a gap between textual similarity and co-visitation suggests a need for re-structuring. This relies on the intuition that visitors should read a number of pages on the same topic, but it is not clear that this is true in the web sites we have considered.

There is a rich literature on search log analysis (see e.g. Silvestri [9] for a recent survey). Our logs represent browse, not search, behaviour, so typical measures such as query reformulations or ranks of clicked results are not directly applicable; however, in future work we hope to align browse and search logs to extract further patterns and explanations.

7. DISCUSSION AND CONCLUSIONS

The trace data collected in web server logs is extremely limited, but does contain signals that we believe are indicative of a user's difficulty navigating a website. Using this to partition sessions into "good" and "struggling" sets lets us find patterns that occur relatively frequently in the latter, and which may help explain users' difficulty.

The algorithms described here are simple, but (as illustrated) are effective in capturing a large proportion of troublesome sessions and providing actionable hints for web analysts. While the particular patterns extracted here will not generalise across sites, the abstract techniques make very few assumptions about site layout, content, or user behaviour and should generalise well. This is comparable to other work on implicit feedback [2], which is generalisable and is easy to use without deep knowledge of a particular site; however, our technique provides more advice on where trouble spots are and what users are thinking.

(Although they cannot generalise across sites, the patterns here do generalise across *time* to some extent. Using data from another week later in the same month, the top five patterns of Section 4.5 still cover 33% of struggling sessions with 15% precision.)

We are building some of this analysis into a tool for web authors. The "web authors' sidebar" analyses server logs and shows the results in the context of a web page, while that page is loaded in a browser (Figure 4). Page transitions or search terms that are significantly more likely in struggling sessions are highlighted with "caution" signs—in this example, there are two transitions that occur more frequently in struggling sessions and might be worth investigating.

By showing this in context, the sidebar can also provide clues as to why people struggle. For example, next pages are labelled according to the referring hyperlink where possible, so authors see what users see, and search terms used in a session provide clues to users' intent.

In future work, we will consider more sophisticated representations of user actions, such as Markov chains [1], which may allow different comparisons between our two sets of users. We are also validating other signals, beyond those described here, which may tell us a visitor is in difficulty. Other records of users' actions may contain further clues, either signals of a struggling session or hints to the cause of confusion, although we are naturally constrained by what we can record at a web server and what users and clients will allow.

We are working with web analysts to check whether these patterns, and the tool in Section 5, do in fact provide useful leads for day-to-day triage and web editing. Although we have not yet carried out a formal evaluation, the tool is in regular use, feedback to date has been good, and analysts have used it to inform changes to their site.

8. ACKNOWLEDGEMENTS

We thank the agency's web analysts for their input, as well as access to their log data. Catherine Wise and Brian Jin contributed to the software described in Section 5.

9. REFERENCES

[1] J. Borges and M. Levene. Data mining of user navigation patterns. In *Proc. Int. WEBKDD'99 Workshop on Web Usage Analysis and User Profiling*, number 1836 in Lecture Notes in Artificial Intelligence, pages 92–112, 1999.

[2] M. Claypool, P. Le, M. Waseda, and D. Brown. Implicit interest indicators. In *Proc. Intelligent User Interfaces*, pages 33–40, 2001.

[3] S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White. Evaluating implicit measures to improve web search. *ACM Trans. Info. Systems*, 23(2):147–168, 2005.

[4] B. J. Jansen. *Understanding user-web interactions via web analytics*. Number 6 in Synthesis lectures on information concepts, retrieval, and services. Morgan & Claypool, 2009.

[5] M. Lassila, T. Pääkkönen, P. Arvola, J. Kekäläinen, and M. Junkkari. Unobtrusive mobile browsing behaviour tracking tool. In *Proc. Information Interaction in Context*, pages 278–281, 2012.

[6] C. Liu, J. Gwizdka, and J. Liu. Helping identify when users find useful documents: Examination of query reformulation intervals. In *Proc. Information Interaction in Context*, pages 215–224, 2010.

[7] N. R. Mabroukeh and C. I. Ezeife. A taxonomy of sequential pattern mining algorithms. *ACM Computing Surveys*, 43(1):3:1–3:41, Dec. 2010.

[8] T. Nakayama, H. Kato, and Y. Yamane. Discovering the gap between web site designers' expectations and users' behaviour. *Computer Networks*, 33:811–822, 2000.

[9] F. Silvestri. *Mining Query Logs: Turning Search Usage Data into Knowledge*, volume 4 of *Foundations and Trends in Information Retrieval*. now Publishers, 2010.

[10] K. L. Smith and P. B. Kantor. User adaptation: Good results from poor systems. In *Proc. SIGIR*, pages 147–154, 2008.

[11] M. Spiliopoulou, C. Pohle, and L. C. Faulstich. Improving the effectiveness of a web site with web usage mining. In *Proc. Int. WEBKDD'99 Workshop on Web Usage Analysis and User Profiling*, number 1836 in Lecture Notes in Artificial Intelligence, pages 142–162, 1999.

[12] P. Thomas, A. O'Neill, and C. Paris. Interaction differences in web search and browse logs. In *Proc. Australasian Document Computing Symposium*, Melbourne, 2010.