THE UNIVERSITY OF
WARWICK

warwickpublicationswrap

highlight your research

http://wrap.warwick.ac.uk/

Human Factors and the WWW: Making sense of URLs

Dominic Stanyer and Rob Procter

Institute for Communicating and Collaborating Systems
Division of Informatics, University of Edinburgh
Scotland
dsjs|rnp@dcs.ed.ac.uk

Abstract: We present a study of how WWW users 'make sense' of URLs. Experiments were used to investigate users' capacity to employ the URL as a surrogate for the resource to which it refers. The results show that users can infer useful information from URLs, but that such improvisation has shortcomings as a navigation aid.

Keywords: Universal Resource Locator (URL), usability, World Wide Web, navigation

1 MOTIVATION

The typical user interface is an abstraction over the behaviour of the underlying system. The user interacts with system functionality through this abstraction which hides the complex realities of the underlying mechanisms (Dourish and Button 1998). We argue that the typical WWW browser hides too much. Users often find themselves having to navigate without a clear idea of either 'where' they are going, or of 'how long' it will take to get there.

Such deficiencies force users to improvise, to use any means to hand that might afford them a window onto a richer picture of the content, form and behaviour of the underlying system. As Nielsen (1999) observes, the Universal Resource Locator (URL) is one such window. In earlier experiments, we explored ways in which URLs may afford useful clues concerning referenced resource content, location and even the behaviour of the download process (Stanyer and Procter 1999). We report here preliminary results from a follow-up study. This is part of on-going research to develop an understanding of how users use the World Wide Web. The goal is to improve WWW usability through better browser interfaces.

2 THE URL

Every WWW resource has a unique reference, or URL, by which it can be retrieved. The three major parts of a URL are `site`, `path` and `resource` name, where the latter two parts being optional. Breaking down the URL into its components may enable users to extract useful information about content, location and download behaviour of the referenced resource. An experiment was undertaken to explore users' ability to make sense of URLs in this way.

3 EXPERIMENTAL METHOD

The subjects were eighteen Edinburgh University students with varying degrees of WWW experience. Each was presented with ten URLs in random order. They consisted of either: `site` name (short), `site` and `path` name, or `site`, `path` and `resource` name (long), and were chosen to represent a range of inference affordability. Subjects were asked to rate their confidence in predicting content, location, and download time on a scale 1 to 6. Downloading was then initiated, and subjects asked to relate the observed behaviour to the URL. On download completion, subjects were asked to rate the accuracy of their predictions. Subjects' remarks were recorded.

> www.dcs.shef.ac.uk/~wmlh/ipcat97book.htm
> "It's at Sheffield University, and it's in someone's personal home directory because it has a tilde for a UNIX account, home directory, and there is some acronym 97 which presumably means that it's the proceedings or something to do with the year 1997, that's what I would assume, I don't know that it will be. Book indicates that it might be an on-line version of a document ... like a paper document maybe and htm indicates that it is hypertext. My intuition would say that book htm could be fairly fast because it could be a big text file representation of a book. Again ... it is just me implying that."

Figure 1: Extract from subject transcript.

4 RESULTS AND DISCUSSION

Averages of subjects' confidence and accuracy in location, content and download time prediction are shown in Table 1. They suggest that subjects were more confident at inferring location (e.g., see Figure 1) than content or download time. Subjects also rated their accuracy of location prediction higher than content or download time. Accuracy is rated quite highly for all predictions. We suspect that this may be an effect of self-assessment. Performance in both tasks was better for long URLs.

The transcripts showed that subjects made sense of URLs in ways that were broadly consistent with URL format and constituent parts, but that they were easily mislead. E.g.: location inferences might be driven by recognition of a site name sub-domain, rather than a 'parsing' of the whole name. E.g.: most subjects based their inference of the location of `uk.samba.org` on its first sub-domain. There was a correlation at the 0.95 level between location and download predictions, suggesting that subjects often based their estimate of the latter on the former. This was confirmed by the transcripts. Other bases for download time inference included inferences about site popularity. E.g.: "Microsoft is always slow because so many people use it". Once download had started, subjects were able to rationalise observed download behaviour by reference to the URL. Subjects' strategy for inferring content had to be that URL parts were meaningfully named. It may be that lower confidence in content predictions reflected

awareness of the unreliability of such an assumption. Content inferences were often conditioned by 'common sense'. E.g.: subjects were very confident about `www.scottishpower.plc.uk`, but much less so about `uk.samba.org`.

5 SUMMARY AND CONCLUSIONS

The URLs were chosen because they exemplified different forms, not because they were typical of what users may encounter 'in the wild'. Nevertheless, the results suggest that though users can make quite artful use of URLs as surrogates for WWW resources, such improvisation has its limitations. It would seem that users are best at inferring location, yet location may be less valuable for navigation than either content or download time (Johnson 1997).

We conclude that browsers should afford users better means for navigating and understanding WWW behaviour. In particular, techniques such as link traversal preview and download time affordances deserve investigation (Kopetzky and Mulhauser 1999, Stanyer and Procter 1999). Such techniques can provide with browsable levels of abstraction without making too much of a commitment as to how these abstractions may actually be used.

REFERENCES

Dourish, P. and Button, G. (1998) On Technomethodology: Foundational Relationships Between Ethnomethodology and System Design. Human Computer Interaction, Vol 13, 395-432.

Johnson, C. (1997) What's the Web Worth? The Impact of Retrieval Delays on the Value of Distributed Information. Workshop on Time and the Web, Staffordshire University, June.

Kopetzky, T. and Mulhauser, M. (1999) Visual preview for link traversal on the WWW. Proc. 8th International WWW Conference, Toronto.

Nielsen, J. (1999) URL as UI, Jacob Nielsen's Alertbox, March.
http://www.useit.com/alertbox/990321.html

Stanyer, D. and Procter, R. (1999) Improving Web Usability with the Link Lens. Proc. 8th International WWW Conference, Toronto.

	Confidence	Accuracy
Location	4.9	5.2
Content	4.2	4.6
Time	3.9	4.0

Table 1. Average subject prediction confidence and accuracy.