Quantum Web Fields And Molecular Meanderings
Visualising Web Visitations

Geoff Ellis
School of Computing and Engineering
University of Huddersfield
Queensgate, Huddersfield, HD1 3DH, UK
g.p.ellis@hud.ac.uk

Alan Dix
Computing Department
Lancaster University
Lancaster, LA1 4YR, UK
alan@hcibook.com

http://www.hcibook.com/alan/papers/avi2004-web-fields/

Abstract
This paper describes two visualisation algorithms that give an impression of current activity on a web site. Both focus on giving a sense of the trail of individual visitors within the web space and showing their navigation paths. Past web activity is used to produce a spatial mapping of pages, which results in highly traversed page links lying close together in the 2D visualisation space. Pages visited by typical individual visitors thus form intelligible paths when plotted in the visualisation space. Both techniques attempt to enhance user awareness and experience, but they differ in their balance between utility and aesthetics.

Categories and Subject Descriptors
H.5.2 User Interfaces – graphical user interfaces, H.5.3 Group and Organization Interfaces – web-based interaction, H.5.4 Hypertext/Hypermedia – navigation, I.2.6 Learning – connectionism and neural nets

General Terms
Human Factors

Keywords
web visualisation, awareness, self-organising map

1. Introduction
In which we reflect on the way we get a sense of human activity in day-to-day life and introduce the issue of awareness of human presence on the web.

In Monte Carlo the manager of a large casino glances at the array of closed circuit TV cameras. The security staff are looking closely for signs of unusual activity, sometimes zooming in on a particular table, but the manager, she simply gets a feel for the busyness of the casino, the games that are most crowded, those that seem unusually quiet.

Sitting in his office, the professor hears the muffled sound of footsteps and hubbub of voices through the window. Without realising it, he stretches and looks at the clock. "Ah it is class change-over time", he realises – he needs to hurry to his lecture. As he starts down the stairwell he stops for a moment to catch sight of the swirling clouds and the trail of the high jets across the sky, then looking down notices the grass below and the worn marks of countless footsteps that eschew the concrete paths laid out for them.

All around us are signs of the activity of others. Some of these we examine explicitly, some merely intrude gently into our background awareness.

Whilst we are used to interpreting signs such as footsteps in the physical world, this is far more difficult in an information space such as the web. As a site owner it is good to have a feel for the people visiting your site, but simple hit counts or logs do not capture the essential humanity of visitors.

In this paper we describe two visualisations that, in different ways, aim to give a site owner or visitor a sense of past or current human activity. We are not aiming to produce visualisations that can be used to analytically understand these visitations, but instead representations that give an awareness of *human presence*.

2. Background and related areas
In which we review some related literature including the system visualising real footsteps that inspired the web visualisations in this paper

In the CSCW world there is a strong strand of work focused on giving users a sense of awareness of other people's activity or of the behaviour of hidden environmental or electronic data. Sometimes this is achieved through computer displays such as the use of video for remote office shares or ticker-tape information sources. Other systems use ambient displays, such as Natalie Jeremijenko's *Dangling String* that displayed network activity.

Typically these visualizations show current activity remotely, but similar techniques can be used to show past activity. Figure 1 shows a recent installation at Lancaster, the Mood Floor, which uses a weight sensitive surface to record footsteps and then leave slowly fading coloured trails on a projected checkerboard [6,12]. This bringing past activity into the current time we call "absent presence" [7].

There is a large literature on web visualization including visitation patterns, this is usually to help the designer analytically understand or evaluate the web site (e.g. Disk Trees [2]). Various techniques have been used to reduce the complexity of viewing paths through a site, such as clustering

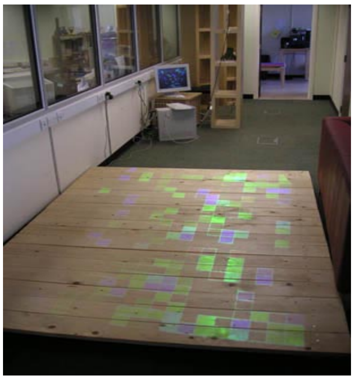


Figure 1. The Mood Floor

[10] and Chi's Dome Tree visualisation [3]. In addition, collaborative filtering techniques based on the actions of past users, have been employed to assist the user navigate a site [1, 14]. However, the Quantum Web Field is less about functional aims and more focused on giving viewers a sense of human presence in the web world.

Several systems directly support awareness on the web. 'Livemaps' [4] shows that annotated site maps can potentially help web surfers to become aware of each other. 'Look who is visiting' [8] makes web publishers aware of the activities taking place at their sites, even when they are engaged in other real-world tasks. In the TOWER environment information spaces, including web sites are laid out in a 3D space where other users are situated near the resources they are accessing [11]. User activity in this information space is also made available through ambient displays, such as a moving fan or small Lego robots called AwareBots, and also through a web accessed Awareness Map [9].

3. Visualising visitations

In which we propose two new visualisations of current web visitation that display in very different ways the paths that visitors are taking through the site.

Imagine a simple display, perhaps a hyperbolic or simple radial layout of web pages on the site arranged in their hierarchy. Each hit generates a small fading dot. This would clearly give a sense of overall activity and also where in the site people were, but no sense of individual visitors. Now colour the dots so that individual sessions have the same colour. While someone is drilling up and down form the home page the coloured dots will move inwards and outwards from the center, but if the site is heavily cross-linked then this intelligible pattern would breakdown and again become more like random dots.

We now describe two systems: the Molecular Meanderer and Quantum Web Fields. These each create 2D visualisation spaces which are laid out so that the 'footsteps' left by visitors in some way look 'human' and so give a sense not just of activity, but also of human presence.

3.1 Molecular Meanderer

The Molecular Meanderer uses a fairly standard ball-and-spring layout, but with a careful choice of similarity matrix in order to create a layout where visitation 'footsteps' are likely to form intelligible paths.

For this kind of visualization there are many fine choices in the algorithm: gravitational inverse square or spring-like linear attractions, level of noise to prevent local minima. However, the most critical thing is the metric used to provide the 'attraction' between nodes. When pages are visited on a path we would like the path to exhibit some locality: in general the pages visited lying close, each to the next one. To encourage this we use a 'comes next' matrix, which measures how often one page is visited immediately before or after another. This is nearly, but not quite a link-followed metric as the 'back' button or history menu will also contribute to 'comes next' metric.

Chalmers et al. [1] used a similar measure but instead of 'comes next' they used co-occurrence within a sliding window of 10 pages. They were interested in using this for automatic recommendations and as navigation aids in the 'open web', hence the concept of clustering this 'occurs close' metric was appropriate. In contrast we are happy for visitation paths to spread out across the visualization space so long as the individual 'steps' are not large.

Figure 2 shows a screen shot from the Molecular Meanderer visualizing the site www.hcibook.com. Each page is a small square and the lines represent frequently traversed links, the red square is the site 'home'. One visitation path is shown numbered and this follows what is clearly a frequent 'linear' path. It turns out that this represents going through the chapter exercise pages in turn.

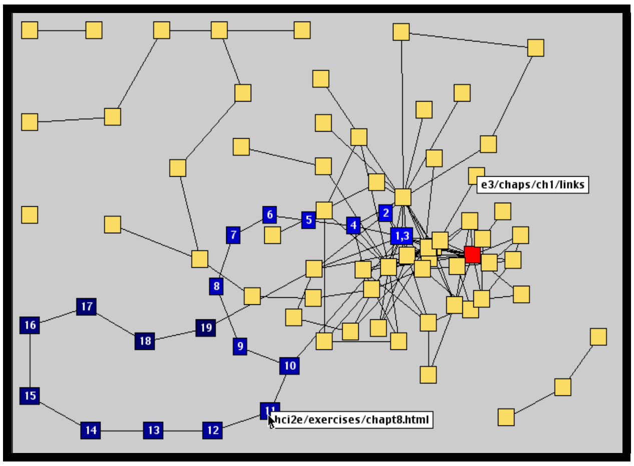


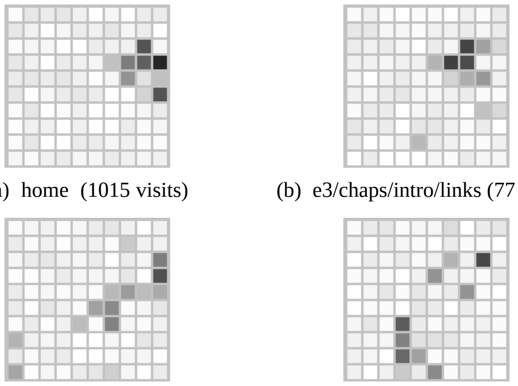
Figure 2. A screen shot from the Molecular Meanderer

3.2 The Quantum Web Field

In the world of large-scale phenomena each object has a unique location in space. However, in the quantum world each particle has a spread out existence – a probability of being at any particular location. Likewise, in the Quantum Web Field, each page is given a probability of being found at each square. These probabilities are arranged so that pages that tend to appear next to one another in session paths have high probabilities on adjacent squares.

When a visit is plotted as a path in the web field, the paths traced are not reproduced exactly and are not deterministic, but are intended to have human-like dynamics. In fact, the human ability to detect 'odd' movements has meant that the Quantum Web Field has also be used to spot non-human activity (web crawlers, site suckers).

Figure 3 shows the web fields for four pages. The darker squares are those where the page has a higher probability of being placed. The algorithm has a ‘crowding factor’ which makes pages spread out over the network of squares, but it does not force each page to occupy a single square. In fact, by adjusting parameters during the procedure that generates the web field, it would be possible to make the algorithm ‘cool’ into a state where there was a single page per square. However, by allowing each page to have a more diffuse mapping to the net we are able to visualise sites where there are more pages than squares. Also the many–to–many relationship between squares and pages allows some freedom when plotting footsteps through the site, thus giving more human-like paths.



(a) home (1015 visits)

(b) e3/chaps/intro/links (77 visits)

(c) hci2e/resource.html (422 visits)

(d) e3/chaps/ch1/links (21 visits)

Figure 3. Web fields for sample pages

Looking more closely at the web fields in Figure 3, we can see that (a) and (b) have more concentrated locations than (c) or (d). Also, note that (a) which is the site home and (c) the resource page both have a high probability associated with the fourth square down on the right hand side (a and c). There are several hundred pages on this site (www.hcibook.com) and only one hundred squares. Finally, note that (d) has a bimodal nature. It is the ‘links’ web page for chapter 1 and one area corresponds to being close to the links area of the ‘intro’ web page (b). The transition between these would be quite common for someone skimming the ‘links’ pages, hence the algorithm places the pages closely on the web field.

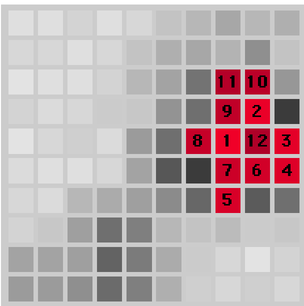


Figure 4. Path through the web net

In the quantum world, particles only get a single location when they are observed. Similarly, when a session visits a page, a single square is coloured as the ‘wave function’ collapses. The location for the page is chosen in a random, yet proportionate manner, based on the probabilities in the page’s web field. The location of subsequent pages in the session also depends on their web field, biased by the distance from the last visited page. The overall effect is that a session path visits a series of squares that tend to be close to one another, but have the occasional jump across the net. Identical paths do not necessarily hit the same squares, although they will tend to follow similar paths across the net. This leads to a visualisation that has enigmatic quality, giving a sense of purposeful activity and human-like variety, but defying a simple explanation.

Figure 4 shows a typical path across the net. To ease interpretation, the squares are numbered with the order in which they are visited and cells also fade with time.

4. Mechanisms

In which we explain the algorithms: using a 'comes nextness' matrix and log sampling.

4.1 Software architecture

Both the Quantum Web Field and the Molecular Meanderer have a similar algorithmic architecture.

(a) web logs are pre-processed to extract sessions and create co-visitation statistics

(b) the comes-next matrix is used to produce a 2D visualisation space

(c) live web visits are sampled

(d) the sampled sessions are drawn into the 2D visualisation space

Figure 5 shows these steps. Steps (a) and (b) are part of a pre-processing stage and steps (c) and (d) are carried out in real time to give the actual visualisation.

Steps (b) and (d) depend on the visualization and have already been described, so we will now expand slightly on (a) and (c).

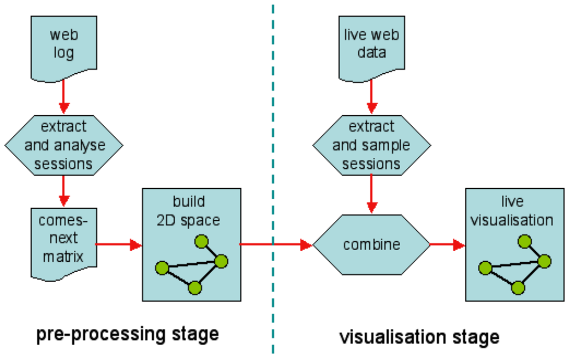


Figure 5. Stages of visitation visualisation

4.2 Log analysis

Although we are considering using specially instrumented sites, for preliminary work we have used standard web logs. We use fairly well trodden data cleaning: removing crawler activity, ignoring images and other non-html files, identifying aliased parts of the site and reducing all urls to id codes. Sessions are detected using the IP address of the web client with a timeout. This is not ideal and an instrumented site could use cookies for more accurate session tracking.

Once the sessions have been extracted it is a simple matter to create the ‘comes next’ matrix. For any pair of pages p_i and p_j, the (i,j) location of the ‘comes next’ matrix is the number of times p_i occurs before p_j plus the number of times p_j comes before p_i.

4.3 Live sampling

We need to sample visitors in order to avoid swamping the visualisation. A previous paper examined the use of sampling in detail [5]. In the case of web visitors the sampling regime can be relatively simple as each visitor is largely 'independent'. This implies that instead of randomly deciding which visitors to display we can simply choose every n-th one where 1/n is our sampling density.

However, if the site receives fairly constant number of visitors this sampling rate can be fixed. However if the site has very quiet and very busy periods then the sampling range may have to b adjusted so that the visualisation does not become either too crowded or empty! The mapping between number of visitors and number of visualized paths needs to be monotonic to give some sense of business, but non-linear to avoid extremes (see fig. 6).

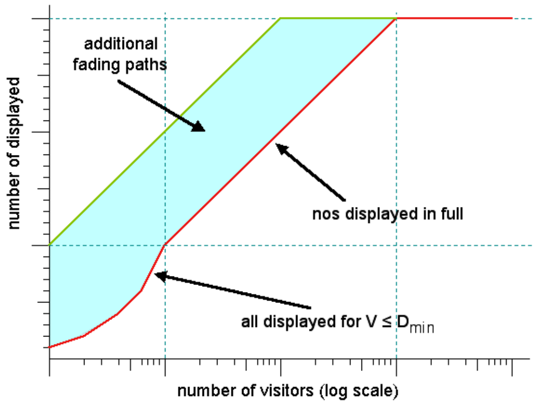


Figure 6. Non-linear sampling function

5. Discussion and future work

In which we contrast the two visualisations and discuss potential future directions.

In the Molecular Meanderer and Quantum Web Field we have seen that logs of past activity can be used to structure non-planar structures (web pages) into spaces that are suited for visualisation and preserve the human-ness of paths. We have also seen that it may be necessary to sample logs of past activity in order to reduce display density. However, this needs to be done in ways that preserve key attributes of the data, in this case the path of a session (leading to session sampling) and the sense of volume of activity (leading to a non-linear sampling regime).

The two visualisations have slightly different uses. The Molecular Meanderer has a clear one-to-one relationship between nodes and pages so that similar paths can be readily identified and where desired, the visualisation can move from being an ambient background display to an investigative foreground activity.

In contrast, the Quantum Web Field is more enigmatic, a form of 'informative art' rather like the Mondrian-like bus time display [13]. The layout is designed more to give a 'feel' of intelligible human presence than to be actually comprehensible.

We intend to install these visualisations on the www.hcibook.com web site shortly so they can become part of the experience of ordinary visitors. Whilst the initial designs are aimed at showing all visitors on one display, we are also considering using the Quantum Web Field as a visualisation of the user's own path through the site.

6. ACKNOWLEDGMENTS

This work was supported by EPSRC grants Tracker and Equator.

7. REFERENCES

[1] Chalmers M., Rodden K. and Brodbeck D.. The order of things: activity-centred information access. In *Proc. 7th Intl. World Wide Web Conference*, (Brisbane, Australia, April 1997). Elsevier, 1997, 359–368.

[2] Chi, E., Pitkow, J., Mackinlay, J., Pirolli, P., Gossweiler, R.and Card, S. Visualizing the evolution of web ecologies. In *Proc. of CHI98*. ACM Press, 1998, 400–407.

[3] Chi E, Pirolli P, Pitkow J, (2000) The Scent of a Site: A System for Analyzing and Predicting Information Scent, Usage, and Usability of a Web Site, In *Proc. of CHI 2000*. ACM Press, 2000, 161–168.

[4] Cohen, D., Jacovi, M., Maarek, Y. and Saroka, V. Livemaps for collection awareness, *Intl Jnl on Human-Computer Studies, 56, 1,* (Jan 2002), 7–23.

[5] Dix A and Ellis G P. (2002) by chance : enhancing interaction with large data sets through statistical sampling. In *Proc. Intl. Working Conf. on Advanced Visual Interfaces, (AVI 2002),* (L'Aquila, Italy, May 2002). ACM Press, 2002, 167–176.

[6] Dix, A. Research on the Scrapheap. UsabilityNews, 22th September 2003. www.usabilitynews.com

[7] Dix, A., Sheridan, J., Lock, S., and Ellis, G. (2004). absenT Presence. In EQUATOR Record and Reuse workshop, (UCL, London, 12-13 February 2004).

[8] Gellersen, H. and Schmidt, A. Look Who's Visiting: Supporting Awareness for Visitors in the Web, Intl Jnl on Human-Computer Studies, 56, 1, (Jan 2002), 25–46.

[9] Gross, T., Wirsam, W. and Graether, W. AwarenessMaps: Visualising Awareness in Shared Workspaces. In Ext. Abstracts of CHI 2003. ACM Press, 2003. 784–785

[10] Hendley R., Drew N., Wood A. and Beale R. Narcissus: Visualizing Information. In *Proc. of the 1995 Information Visualization Symp*., (Atlanta, GA). 1995, 90–96.

[11] Prinz, W. and Gross, T. Ubiquitous Awareness of Cooperative Activities in a Theatre of Work, In *Proc. of Fachtagung Arbeitsplatzcomputer: Pervasive Ubiquitous Computing* (Oct. 10-12, Munich, Germany) VDE Publisher, Berlin, Germany, 2001, 135–144

[12] Scrapheap (Computing) Challenge. [online] http://polo.lancs.ac.uk/scrapheap. (Accessed Dec. 2003).

[13] Skog, T., Ljungblad, S. and Holmquist, L.E. (2003) Between aesthetics and utility: designing ambient information visualizations, In *Proc. of InfoVis 2003*, (Seattle, Washington), 2003.

[14] Wasfi, A. Collecting User Access Patterns for Building User Profiles and Collaborative Filtering. In *Proc. of IUI'99*, Los Angeles, ACM Press, 1999, 57-64