Measuring the readership of a health-related website

Briggs JS

Healthcare Computing Group, University of Portsmouth, United Kingdom

Correspondence to: Dr J.S. Briggs, School of Computer Science and Mathematics, University of Portsmouth, Milton Campus, Locksway Road, Southsea, PO4 8JF, UK

Tel: +44 23 9284 4281; Fax: +44 23 9284 4006; Email: Jim.Briggs@port.ac.uk

Abstract

The UK National Database of Telemedicine (NDTM) (http://www.dis.port.ac.uk/ndtm) has been set up with the sponsorship of the British Government to provide a background source of information to anyone researching the field or proposing a trial or a larger scale implementation of Telemedicine.

Publishers in any media are always conscious of the need to know who their readership is. In traditional publishing this is done by reader surveys and other similar means. In electronic publishing, the technology often provides other ways of eliciting this information.

Conventional means of measuring the access to an Internet website involve the use of counters. These are limited in that they only tell you how many people have accessed a particular page - they do not tell you anything about those people. Analysers that look at the access logs of a website can tell you a little more, such as the name associated with the Internet address used by a reader. However, more information than this can be gleaned by seeking the co-operation of the reader's web browser.

This paper describes how we used such a means to collect data about access to the NDTM website, and our analysis of the readership.

Introduction

Historical background to the NDTM website

In 1996 the Research and Development Division of the United Kingdom (UK) Department of Health (RDD) commissioned a survey of telemedicine activity in the UK[1]. The aim of the survey was to provide information that could help with the formulation of policy in this area and to identify the barriers to progress. A further survey of telecare applications was conducted in 1997[2].

In 1998, it was decided that the information collected in these surveys needed to be disseminated more widely to the UK telemedicine community. A website was suggested as a suitable means of doing so and the University of Portsmouth was commissioned first of all to assess its feasibility and then to launch an initial version.

The website is known as the **UK National Database of Telemedicine** (NDTM)[3]. As well as a list of telemedicine and telecare projects, it also contains a list of telemedicine equipment and service suppliers and references to many other sources of information[4]. It is targeted at anyone researching the field or proposing a trial or a larger scale implementation of telemedicine.

NDTM was officially launched on 27th October 1998 and publicised extensively by electronic means such as email and newsgroup messages. Before and after its launch, it has been kept up to date with information about new projects and modifications to the details kept about existing ones. The database currently (August 1999) contains information about 140 projects and companies.

The University of Portsmouth still hosts and maintains the website on behalf of the Department of Health, but now with the support of the British Library (Telemedicine Information Service). A three-year contract for continued maintenance and updating of NDTM commenced in May 1999.

Assessing the readership of the website

Most publishers monitor the readership of their publications by one means or another. In the publishing industry, it is common to assess the popularity of particular authors or material, and to judge the success or otherwise of marketing activities. For printed media, sales are the main measure used. With magazines and newspapers, conventional market research activities such as surveys can help deduce how many people read each copy sold. For electronic publication, particularly online on the World-Wide Web (WWW), different measures need to be used. The measure normally adopted is to count the number of readers of a particular web page or site.

In the case of NDTM, we wanted to monitor its readership for a number of reasons:

1. To see whether it was creating a community of telemedicine practitioners and potential users. This was one of the Department of Health's objectives for the project since they wanted a group with which they could consult informally on policy and technical issues.

2. We may want to tailor content to the particular audience it is reaching.

3. We want to know who has seen our publicity and read the website, and thereby to identify groups of people to whose attention we need to bring it.

4. There would be no point in expending resources on keeping it up to date if nobody was reading it.

While it is possible to do a variety of surveys in different forms to determine the demographics of NDTM's readership, in this paper we concentrate exclusively on **technical mechanisms based on the website itself**. We describe the mechanism we use for tracking usage of the website and show that it provides more information than comparable approaches. We also discuss the information we have collected and draw some conclusions from that data.

Methods

Anyone surfing the Internet for any time would have noticed two common ways by which web page authors measure their readership: guestbooks and counters. These are both visible to the reader. Those people who have set up their own web servers will know that there is another way, hidden to the reader, by configuring the web server software to keep logs of accesses to it. In the next four sections, we discuss the pros and cons of these approaches and then go on to discuss the technique we used for NDTM.

Guestbooks

Guestbooks are normally web forms that readers can complete with comments about a web page or web site. A number of free guestbook services are available on the Internet[5]. These provide the necessary forms and the database in which to store the comments. They usually also provide a mechanism to view the comments in the guestbook - again in the form of a web page.

Guestbooks are not a reliable means of assessing readership of a web site. There is no compulsion on a reader to complete an entry, and there is no evidence to show that the readers who do record comments are representative of the majority.

Counters

A web page counter is usually an image of a number. The image is created dynamically each time it is accessed; the number shown being incremented each time. The counter image is often created on a separate server (a large number of organisations offer to do this for free[6]) on which the count is stored and incremented.

Counters are a moderately reliable way of counting the number of readers of a particular web page or site. They will not however count all readers – specifically, two categories will not be counted:

- those readers who have set their browser not to automatically load images when a page is read (if the browser does not request the counter image, the act of reading the page will not be counted)
- those who abort the loading of a page before the counter image is requested

The other disadvantage of a counter is that it tells you nothing about your readers other than how many of them there have been.

Access logs

Most web server software can be configured to retain a log of accesses to it. In the most common cases, every time a web page is requested by a browser the server records:

- the Uniform Resource Locator (URL) of the page requested

- the Internet address of the requesting browser
- the time and date the request occurred

Analysis of this log can then be made to report on such things as the most frequently accessed pages, the Internet addresses generating the most requests, and the times of peaks and troughs in demand.

Access logs are more reliable than counters in that they count the requests for a page rather than that of some element within it. It is therefore not possible to bypass the counting, meaning the results are more accurate. However the price paid for this accuracy is the need to store the logs on the server machine. For a website that is attracting large numbers of accesses, the size of the log can grow very quickly. To do long-term analysis of readership may therefore require a large amount of disk space.

Since the access log is server based, it cannot tell you how the reader came to request a particular page. It would be useful to know what page the reader had read immediately prior to reading your page, since this probably indicates a page that contains a referring link. To do this sort of more sophisticated analysis requires the co-operation of the browser. In the next section we shall see how this can be done.

A more sophisticated approach

The approach that we adopted was to subscribe to a free Internet service called *Extreme Tracking*[7]. To make the technique work, we had to include a few lines of JavaScript code (provided by Extreme Tracking) into the web page we wanted to measure – in our case the NDTM home page. This code requests an image from the Extreme Tracking server, but in so doing transmits information including details of the reader's browser and the URL that the browser accessed prior to requesting the current page (the so-called *referrer address*).

This approach has two advantages in comparison with conventional server access logs:

1. It is Extreme Tracking who are storing the access details so the approach makes no demands on disk space on your server. (Extreme Tracking's reports analyse a year's worth of accesses.)

2. The referrer address (where available) is stored with each access.

The technique has one disadvantage and that is the same one as with counters. If the reader's browser has automatic loading of images disabled, then the access will not be counted.

Results

We enabled tracking on the NDTM website some time before it was officially launched. Hence we have been able to monitor the effect that the launch and its attendant publicity had on readership. We have also been able to draw some conclusions from the data that has been provided to us.

Since launch there have been over 9000 visits to the website from over 5000 different Internet addresses. The peak was in the week after launch when it had 400 visitors in 4 days. Since then the number of accesses has tailed off slightly, but nine months later NDTM is still attracting of the order of 800 visits per month, indicating the site's popularity and usefulness. Most accesses are on weekdays and during normal UK working hours. Taken together with the fact that the Internet domain that generates most visits is the ".uk" one, we conclude that

most readers are based in the UK and accessing the site as part of their jobs rather than as a spare time activity. This is encouraging for the take-up of telemedicine in the UK.

Analysis of referrers shows that the majority of visits to NDTM are via links from other websites. Of the 288 different websites identified, the one operated by the UK National Health Service (NHS) Information Authority generates the most visits, suggesting that many readers are from within the NHS. The second most common form of referral is from search engines. NDTM appears to be known by 16 of them, however Yahoo[8] generates far more references than all the other ones put together.

Readers of this paper interested in seeing up-to-date reports on NDTM's readership may look at the following URL: http://extreme-dm.com/s/?tag=ndtm.

Discussion

The results we have obtained from using Extreme Tracking have led us to a number of interesting observations about NDTM's readership. However, there are a number of questions that we still wish to answer and issues that we need to address:

1. The figures we have show the number of unique Internet addresses that have accessed NDTM. This cannot equate precisely to the number of readers since:

    a) Several readers may share the same computer and hence the same Internet address.

    b) Readers who access NDTM through an Internet Service Provider (ISP) that allocates addresses dynamically probably get a different address each time they access the site.

    We thus do not have a precise value for the number distinct readers, nor how many visits are repeat visits.

2. To assess the impact that NDTM has on telemedicine in the UK, it would be useful to know how many of our readers fall in to each of two categories: users (or potential users) of telemedicine systems (e.g. healthcare professionals) and suppliers of telemedicine equipment and services. We hypothesise that further analysis of the Internet addresses of NDTM's readers might go some way to answer that question. For example, readers in the ".nhs.uk" domain are probably users, while those in the ".co.uk" and ".ltd.uk" domains are probably suppliers. While Extreme Tracking cannot answer this question, we should be able to tell this from server access logs.

3. What effect will further publicity have on the level of readership? We have not issued any further publicity of NDTM's existence since its launch. We would like to conduct a series of experiments to ascertain the effect that targeted publicity has on access. The first of these was to present this paper to a largely Canadian audience at the Congress on Telehealth and Multimedia Technologies in Edmonton. We look forward to seeing an increase in the number of accesses from the ".ca" domain in the following weeks.

4. How many potential readers of NDTM fail to find the website? We plan to regularly test the efficacy of the Internet search engines by giving them possible search strings and seeing whether they suggest NDTM in response. The results can be used to suggest steps to take to improve efficacy in the future, e.g. by registering NDTM with new search engines or by improving the quality of keywords associated with our pages. We are particularly interested in ensuring that NDTM's contents will be found by healthcare-specific search engines and other online specialist library facilities.

Acknowledgements

References

1   Curry RG. Telemedical Activity in the United Kingdom: A Review and Assessment, September 1996 [unpublished report]

2   Curry RG. *A Review and Assessment of Telecare Activity in the UK and Recommendations for Development*, September 1997 [unpublished report]

3   National Database of Telemedicine, Portsmouth, UK, URL: http://www.dis.port.ac.uk/ndtm

4   Briggs JS, Hart P, Norris AC, and Curry RG. "The UK National Database of Telemedicine website". In Arvanitis TN, Keevil SF and Woodall J, editors, *Proceedings of MEDNET 98* (3rd Annual World Congress on the Internet in Medicine); 1998 Nov 16-19; London. Birmingham: University of Birmingham; 1998. p. 33, [ISSN 1463-9394].

5   "FREE Guestbooks for your Web Site", http://www.thefreesite.com/guestresource.htm

6   "FREE Web site counters & trackers", http://www.thefreesite.com/freecounters.htm

7   Extreme Tracking, URL: http://www.extreme-dm.com/tracking

8   Yahoo, URL: http://www.yahoo.com