Abstract

Nowadays, for many people, the World Wide Web (WWW) is the first place to go to look something up, to find that bit of information. However, even though people have their favourite sites, and their favourite search engines, they often seem to miss that bit of information. This could very well be because it is hiding on a small, unpopular, enthusiast's Personal Home Page. The author believes that there is more information on the Web than that which one will find on the major, "commercial-style" sites. Hence, this paper looks at the possibility of using Personal Home Pages (PHP) as an information resource, not only for the academic, but the web-surfing world in general.

Keywords

Personal Home Pages, Personal Web-pages, Online resources, Web resources

Introduction

For as long as the World Wide Web has been around, it has continued to change and evolve. To begin with, it was simply a medium used to store and distribute information cheaply. Nowadays, it has a multitude of uses including numerous forms of entertainment (e.g. online movies, gaming) and business (e.g. online selling, business systems integration) among others. However, it is its role as the world's largest information resource that we are concerned with it here. The article proposes the evaluation of the worth of the Personal Home Page (PHP), possibly an underrated, but very useful, information resource.

The PHP is the subject of much abuse. It has been said that a PHP is a website that has been created in bad taste, with lots of colour clashes, a self-promoting style of writing (lots of I's and me's), with numerous grammatical mistakes, an unusually long URL, links to several incomplete pages, the odd tasteless animation and an amateur photograph (IRON, 2003).

This probably explains why the PHP, even though it made its appearance relatively early on in the World Wide Web (as early as 1993 (Koch, 1993)) has not merited the attention of the research community quite as much as other characteristics of the Web.
What is a personal home page

The PHP as a subject of research concerns many disciplines. These include computing, information science, psychology and even journalism, to name but a few. As a result, perhaps, it is not too surprising that there is more than one definition of the PHP doing the rounds.

De Saint-Georges (1997) provided a tentative definition of the PHP as a "presentation of the self in digital (hypertextual) form, authored by one individual, and which (i) emphasizes a person (minimally, by a picture or a name); and/or (ii) a person's current activities; and/or (iii) professional experience; and/or (iv) displays a person's interest (in the body of the text and/or through hyperlinks to other sites)."

Clearly then, recognising something as a PHP requires that the site has to:

1. Claim that it was a PHP
2. Have personal information on it (such as a CV, photograph etc.)
3. Represent a person rather than a group

The Campus-Wide Information Service (CWIS) website at Murdoch University, Perth in Western Australia (1999) offers another, somewhat fuller definition: "A HTML document prepared by an individual, that they describe as their Home Page. The document may be held on a personal computer at Murdoch, or made available from the individual's home directory area on a host computer at Murdoch, or made available from a computer accessible via the Internet on which the individual maintains a user's account. A Personal Home Page may include personal and biographical details and may offer a link to a document maintained for that individual on a CWIS Server. A Personal Home Page that is accessible via a CWIS Server will include a link to the Disclaimer Notice maintained on the CWIS Host Server."

Buten's First World Wide Web Personal Home Page Survey (1996) provided two rules (about the contents of the web-page) to which a web-page must adhere in order to be included in its survey (there were other rules about location, but those are beyond the scope of this paper). Basically, a web-page was considered a PHP if it was referred to as such (self-defining in other words), or if the page was listed under a personal name (e.g. Jack's page).

Weaver (2000) also conducts a survey, in an attempt to determine whether the viewing of personal web-pages are part of a reference librarian's duties. For the purposes of the survey, PHPs are given a definition as pages which are "wholly under the control of individuals, and not functioning as official library pages".

Dillon and Gushrowski (2000) conducted a PHP study in which they examined over one hundred PHPs and deduced a number of typical elements that can be found within these. Included in the list of elements were very common generic web-page ingredients such as a 'Title' and 'Create date', and then more PHP specific ones such as 'Brief bio' and 'Guestbook'. In the study, these elements were passed on to information science students who put together lists of elements, which they expected to find on PHPs. From the definitional point-of-view, the most interesting aspect is where these hundred PHPs were found. They were just picked up from two 'home page resources', namely The PeoplePlace (http://www.tiac.net/users/domvon - now deactivated) and Personal Pages Worldwide (http://www.utexas.edu/world/personal). The research was conducted with just an assumed definition, not an explicit one.
Papacharissi (2002a) also conducted a study involving PHPs and once again the location from which the pages were identified from says the most in terms of a definition. One thousand PHPs were simply picked up from four 'Personal Home page Providers'. These were Yahoo! Geocities (http://geocities.yahoo.com/), AOL Hometown (http://hometown.aol.com), MSN Homepages (http://homepages.msn.com) and Earthlink (http://start.EarthLink.net/search). Papacharissi (2002a) took one extra step in that the pages that were 'affiliated with or constructed by a commercial organisation or other institution' were ignored. However, once again, the research was conducted with no explicit definition.

For the purposes of this article we have adopted the definition of de Saint-Georges (1997). A web-page is considered a PHP if it is thought to have been authored by one person with the purpose of presenting that person's interest and persona.

**Aims and objectives**

The aims and objectives for this study as a whole are threefold. Firstly, to discover the degree to which PHPs make an informational contribution to the World Wide Web as a resource. Secondly, to provide a better understanding of what PHPs on the Web have to offer in terms of types of information vis-à-vis more traditional information sources on the Web. Thirdly, to provide more accurate definitions for PHPs and other PHP related web-pages. As such, this paper acts as an introductory inspection by looking at related works, with a view to meeting the aims and objectives mentioned.

**Literature review**

Dillon and Gushrowski (2000) conducted a study of 100 PHPs and a list of elements found on those web-pages (e.g. title by frequency of occurrence, email address by frequency of occurrence) was recorded. Eight experimental pages were then created. These pages were then shown to 57 students who ranked them from 1 to 8. The students were then asked which elements a PHP should have. The results suggest that the PHP is now a recognisable media form, and according to the authors, can be considered the "first truly digital genre".

De Saint-Georges (1997) looked at the deictic elements of the grammar found in PHPs. The author took the PHPs of 38 Georgetown University students, and examined the spatial, personal and temporal deixes as found in the texts of the PHPs. The research went on to study the characteristics and patterns of the texts. The result is that the patterns create a certain style of writing (distinct from other styles of writing) which is perceived by the readers as the writing on a PHP. The paper then ends with a tentative definition of what the PHP genre is, based on all these studies.

Papacharissi (2002a) carried out a study, which sought to identify characteristics and attributes of PHPs by examining the 'presentation of the self in virtual life'. A sample of one thousand PHPs was taken from four Personal Home Page service providers (Yahoo! Geocities, AOL Hometown, MSN Homepages and EarthLink Homepages). The results seem to point to PHP authors creating an 'online portrait' of themselves using the tools they have at their disposal (such as guestbooks, links etc.). However, Papacharissi does say that these are not conclusive and more research needs to be done on the subject.

Papacharissi (2002b) carried out another study that looked at the "self online" by examining the purposes that PHPs serve for their authors. In order to do this the web-page contents were analysed, and a survey was carried out. Once again, 1000 pages were taken from Yahoo! Geocities, AOL Hometown, MSN Homepages and EarthLink Homepages, and this time their authors were emailed. This study has sections on PHP author motives, predictors of web-page characteristics, unwillingness to communicate (i.e. authors like or
dislike of face-to-face contact) and contextual age. Among the numerous findings, the results show that PHP authors tend to host pages primarily for information and entertainment purposes. Other reasons (in order of popularity) were self-expression, communication with friends and family, professional advancement and finally, passing time.

Dominick (1999) examined the PHP in regard to the opportunity it gave to the Web author to become a mass communicator and project himself/herself to the rest of the world (however he/she may see fit). As such, the study took 319 PHPs (originally 500 pages) and examined them to find out what web-page authors were doing with the opportunity given to them. Among other things, it looked to see the differences between real self-presentation and 'virtual' self-presentation. It did this by examining the demographics and the contents of the pages. A (very brief) conclusion from the results of this research would show that the PHP is a tool of self-expression, which is used by authors who are primarily, young and male, who create fairly "thin" pages which are tailored to a specific audience. The paper itself has among the results several breakdowns of the contents of the web-pages, including frequency of content items, gender differences in personal information, and gender differences in self-presentation strategy.

Buten (1996) created the Personal Home Page Institute (http://www.asc.upenn.edu/usr/sbuten/phpi.htm), which carried out what was the First World Wide Web Personal Home Page Survey. In this survey, the authors of randomly selected PHPs in the state of Pennsylvania were contacted. A 35-question survey was then sent to the authors, and 151 of 316 mailed surveys were returned. Of those, 121 were usable yielding a 36% return rate. The survey supplied results in many categories such as Population Estimates, Demographics, Internet Use, The Practice of Web Authorship and Webette: Misrepresentation of web-pages.

Bates and Lu (1997) conducted a study of 114 personal home pages, which they took from the People Page Directory (http://www.peoplepage.com). The primary purpose was to analyse the content in the hope of detecting trends or patterns in the design or the content. Specific aspects which were looked at included the purpose of home pages, the degree to which they are self-revealing (in personal terms), their structure and their physical features. The study found that even though certain elements and design features were present often on sites, there was no one feature which was ever-present.

Crowston and Williams (2000) conducted a study of 100 randomly chosen web-pages, and then proceeded to categorise them by genre type (e.g. home page, FAQ, index). The study found that though there were many genres, which already existed in the non-Web world, such as essays and newsletters, there were a few which were completely new, such as the hotlist and the Personal Home Page. The idea behind the research is, since genres make communications more easily understood, creators of web-pages should take advantage of this, and try to use recognisable genres to make pages easier to understand.

Lawrence and Giles (1999) carried out a study into the size of the World Wide Web, search engine coverage and the accessibility of information on the Web. Random IP addresses were tested (out of a total of 4.3 billion possible), and using the few of the tests that actually locate a server, an estimate was made on how many web-pages there are on the WWW. The study then sought to determine search engine coverage, by analysing the responses of these engines to real world queries. With the figures of the size of the Web, and the relative coverage of the search engines, the study determined how accessible web-pages there were. The paper produced, among other things, several tables, such as the distribution of information on the Web in categories (e.g. Health, Personal) and the number of pages indexed by AltaVista and Northern Light (http://www.northernlight.com). According to the research, 83% of Web servers in 1999 contained commercial content,
with 6% scientific/educational and the remaining 11% being shared by personal, government, health, pornography, community, religion and society content. Another interesting outcome of the research is the fact that all search engines combined covered only about a third of the Web.

And finally, Rubio (1996) has written an article about Home Pages, what they are and what they mean. It is an informal article, which talks about home page related topics, such as how hard it can be to describe what a home page is, the relationship between the home page authors and the Web, and also the relationship between home page authors and other home page authors. It attempts to give an insight to the people behind the home pages. It laments the fact that on one side of the world, people make such an issue about their "virtual homes" whilst on the other side, people do not have homes of any sort. It goes further to say that irony on home pages is also prevalent in the writings of the authors, and that often, the worst of the Web gets plenty of attention, simply for being the worst.

**PHP numbers**

It is very hard to tell exactly how many PHPs exist on the World Wide Web. Papacharissi (2002b) points out that there is no directory of personal homepages. However, the number must be very large. In 1996, Buten (1996) estimated that there were 600,000 PHPs in the United States alone. A few years later, in a study conducted by Lawrence and Giles (1999), there were 800 million publicly indexable web-pages of which 2% were personal homepages, putting the number of PHPs at just over 16 million. However, even that number is now a gross underestimate. In December 2000, Wired magazine (2000) reported that Geocities alone had 5.5 million members.

Other indicators also point to the existence of many PHPs. The OCLC Web Characterisation Project has a list of the "Top 50 most frequently linked" websites in 2001. In that Top 50, there are 7 sites which are primarily PHP hosting sites, 3 of which are in the top 10 (OCLC, 2002).

<table>
<thead>
<tr>
<th>name</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geocities</td>
<td>4</td>
</tr>
<tr>
<td>AOL (Hometown)</td>
<td>5</td>
</tr>
<tr>
<td>Tripod</td>
<td>10</td>
</tr>
<tr>
<td>Angelfire</td>
<td>12</td>
</tr>
<tr>
<td>Xoom.com</td>
<td>21</td>
</tr>
<tr>
<td>Earthlink</td>
<td>23</td>
</tr>
<tr>
<td>Compuserve (Our World)</td>
<td>38</td>
</tr>
</tbody>
</table>

**The value of PHPs**

There are several reasons why there is no organised utilisation of PHPs as an information resource. A key reason might be the lack of consistency in terms of content. As there is no traditional incentive (e.g. financial or moral) behind the creation and maintenance of the majority of PHPs, there is no reason to believe they will provide timely and accurate information that the average Web user is looking for. Where the BBC website (http://www.bbc.co.uk) might provide information on which team a football player plays for, a PHP might not be a reliable source of that information. If a PHP did have this information in 1998, the player might have moved on, or in today's climate the team might
be extinct. However, the strength of the PHPs might lie in providing static or relatively static information. As an example, a PHP might provide instructions on how to tune a piano. This kind of information is not heavily dependent on any kind of external factors, such as whether a football team still exists or not, or even time for all practical purposes.

In this light, the following section describes what the author believes to be the stronger and weaker points of information presentation by PHPs.

**Advantages of personal homepages**

The main advantage of the PHP over more commercial-style sites (i.e. non-PHPs such as major news providers and company sites, e.g. CNN and Yahoo!) on the Web could lie in the type and depth of information it carries. Whereas the BBC and CNN (http://www.cnn.com) websites have huge amounts of information, which is both accurate and up-to-date (for the greater part), many pieces of information cannot be found within them. Particularly in topics where commercial rewards cannot or have not yet been reaped, the author believes information is more likely to be found on PHPs of enthusiasts, or solely on PHPs of enthusiasts. An example of such a topic might be the lyrics to the Armenian National Anthem. A search on Google (carried out on 14/8/02, search term: +"armenian national anthem" +lyrics) provided 4 unique hits. Of the four, only one actually had the lyrics to the Armenian National Anthem, and this was on a Geocities site, built by an enthusiast. Perhaps the lyrics to the anthem could be found on other more commercial-style or organisational pages, perhaps on the website of an Armenian Embassy. However, even if this is the case, one could argue that the value of the PHP as an information resource lies in the fact that they are simpler then larger sites to navigate, and at times, simpler to find (as none of these Armenian Embassy sites were revealed by Google).

There are also other information related advantages to PHPs. One of these is the ability to contact the author. Though no research seems to have been carried out specifically on this issue, there has been some research to suggest that PHP authors are more than happy to answer questions about their web-pages. According to Buten’s survey (1996), 34% of PHP authors created their pages to "distribute information to people I don't know with similar interests". According to the same survey, 58% receive emails related to their pages at least once a week, with 45% claiming to talk about their web-pages at least once a week. Furthermore, the survey claims that email is the PHP author’s most popular Internet activity, with 44% sending or receiving more than 10 emails a day. There is more evidence that PHP authors do not mind sending emails about their web-pages by the fact that the survey itself was carried out by sending emails to PHP authors. Papacharissi (2002b) who also carried out a survey, had a response rate of 30%, and Buten (1996) had a very similar 29%.

**Disadvantages of personal homepages**

Perhaps the single most disheartening reason for not wanting to research PHPs is the fact that so many of them seem to be incomplete. Many times a search engine will show a result which appears to be a PHP with the desired information, only to turn out to be an incomplete site or even a dead link. The sheer hopelessness that accompanies such discoveries often leads to the belief that too many of the PHPs are either incomplete or "dead" for them to be worth investigating.

Unfortunately, the problems do not end there. Even if the page is available, complete and functioning, there can be other problems. PHPs can be hard to understand, and poorly organised. They can be completely out of date, as there is no pressure on the authors to keep them up-to-date. Still even if they are up-to-date, the information on the pages can be very biased and sometimes even plain wrong. Finally, the last problem that PHPs face is
the simple fact that they can be complete, organised, up-to-date, and accurate but not be registered on a single search engine (rendering them almost completely inaccessible and hence useless as an information resource to the general public). Perhaps though, this is exactly the reason why more research should go into the topic, to decipher exactly how many such pages exist, and what should be done to show them to the world.

### Table 2. Possible advantages of PHPs against "commercial-style" sites.

<table>
<thead>
<tr>
<th>advantages of php(s)</th>
<th>Advantages of &quot;commercial-style&quot; sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information about non-commercial and unpopular topics</td>
<td>More accurate</td>
</tr>
<tr>
<td>Information in very great depth</td>
<td>More up-to-date</td>
</tr>
<tr>
<td>Ability to contact the author</td>
<td>More complete</td>
</tr>
<tr>
<td>Easier to navigate (due to smaller size)</td>
<td>More accessible</td>
</tr>
</tbody>
</table>

**Use of PHPs as an information resource**

As a separate communication genre and as one of the first completely unique digital genres (Dillon and Gushrowski, 2000), the PHP is an important way to communicate ideas socially. It is also said to be very important for cybercommunities and cyberenthusiasts (Rubio, 1996). By posting messages, making a statement and eliciting responses, people have a level of interactivity which had been missing before the PHP. The PHP gives one the ability to make a statement of almost any size, and keep it online for as long as they wish. Generally, there are quite a variety of uses and roles for the PHP. However, this paper focuses mostly on their use as an information resource.

Buten’s (1996) survey states that the top three reasons for creating a PHP are for "means of expression" at 49%, "learn/practice HTML" at 48% and "distribute information to friends/people" at 43%. This means that just under one in two people use or intend to use PHPs to distribute information. Papacharissi’s (2002b) survey supports this notion, with 33.8% of respondents creating pages which focus on their general interests (this being the most popular response). Meanwhile, on their web-pages (Buten, 1996), 50% of respondents wrote about entertainment, 23% wrote about their research and 19% wrote about sports.

Another interesting statistic from Buten’s survey (1996) is in the expected audiences. The survey has split up the PHPs into commercial and educational samples (according to whether the PHP is found on a .edu site or a .com/.net site). In both samples, 63% of respondents thought that "browsers" (i.e. surfers) would visit their site. Also, both samples thought that "fellow enthusiasts for a topic/hobby" would visit, 61% in the commercial sample, and 42% in the educational sample.

Perhaps the strongest support for the idea that PHPs are useful as an information resource comes from the study conducted by Dominick (1999). From a sample of 319 PHPs, Dominick found that 75% contained information about either "likes" or "dislikes". Breaking that down, it was discovered that 44% (of the 75%) of pages contained information about hobbies, followed by music at 34%, sports at 20%, arts at 11% and books at 8%. This means that at least 239 out of the 319 pages contain the kind of content that would make them useful as an information resource. However, this number is slightly misleading, as the original size of the sample was 500. This number was then reduced by 149 unavailable PHPs and 32 commercial pages disguised as PHPs. This puts the new figure closer to around 48% (rather than the original 75%). However, the accuracy of this figure cannot be treated as definitive either, as it is the result of only one study.
Table 3. Expression of Likes/Dislikes on PHPs from Dominick's study (1999), Sample size: 319

<table>
<thead>
<tr>
<th>Likes/Dislikes on PHPs</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hobbies</td>
<td>44%</td>
</tr>
<tr>
<td>Music</td>
<td>34%</td>
</tr>
<tr>
<td>Sport</td>
<td>20%</td>
</tr>
<tr>
<td>Arts</td>
<td>11%</td>
</tr>
<tr>
<td>Books</td>
<td>8%</td>
</tr>
</tbody>
</table>

Finally, one more reason which supports the existence of unique information on PHPs might be the fact that it is now so cheap to publish on the Web. All one needs to publish information on the Web is access to a computer with access to the Web. This leads people who have no other means to publish with, to be heard through the Web. Whereas people in the past did not have a medium to publish freely available information, they can now. Dominick (1999) also supports the idea, by stating that before the Web, only certain groups of people had access to a mass audience, such as politicians, celebrities, advertisers and media magnates.

All this seems to indicate that there is reason to believe that there is "useful" information to be found on these pages. That is, if the page can be located. According to Lawrence and Giles (1999), search engines are more likely to find web-pages if they are popular. There are two important points here. The first is that, the probability that a search engine will pick up a PHP rises with the number of links that a PHP has pointing to it. The second is that, search engines are having difficulty indexing the entire Web. This basically means that, increasingly larger, more "popular" commercial-style sites will be picked up by the search engines at the expense of PHPs. Perhaps, once the exact "usefulness" of PHPs is determined (if this can be done at all), this bias can be reversed. Alternatively, it might be the case that this bias is justified, by the "uselessness" of PHPs. At any rate, unless there is a closer investigation of the information found on PHPs, this will not be established.

**Conclusion**

There does seem to be a general consensus within the academic world that PHPs are not worth much in terms of information resources. However, people who use the Web on a regular basis seem to agree that every now and then, one comes across a surprisingly useful PHP. At present, it is still unclear as to what kind of role PHPs can play in terms of providing useful information. Indeed, it is unclear what kind of research is needed to determine a role. Regardless though, there is a notable amount of evidence suggesting that valuable information is available on these pages. The question seems to be less "is there any useful information on PHPs?" but more "exactly how useful is the information on PHPs?". Accuracy, reliability and other general quality attributes of PHPs need to be quantified. Likewise, the number of PHPs in existence must be quantified, particularly those that contain what might be "useful information". It might be the case that certain fields of interest have an unexpectedly high number of PHPs. There may be subjects where PHPs are the primary source of information for those with an active interest, or other subjects where the high number of PHPs significantly changes an aspect of this subject, significantly increasing the popularity of that field.

There are many ways in which PHPs might be useful and indeed affect their fields. However, even if other positive effects of the PHP are overlooked, and only the perspective of the web-surfer is taken into account, it might still very well be of significant value to know what is out there on PHPs.
Taking into account the amount of evidence suggesting that at times, PHPs hold very useful information, this course of study is certainly worth pursuing further.

**References**


**Footnote**

Deictic elements in grammar are words such as 'here', 'there' (in place deixis), or 'now', 'then' (in time deixis). There are words which need to be in context for the reader to understand their meaning.

**Bibliographic information of this paper for citing:**

This article has been cited by other articles.

Copyright © 2004, Shant Narsesian.