Feeds as Methodological Tools for Web Research

Andreas P. Giannakoulopoulos
New Technologies Laboratory
Faculty of Communication and Media Studies - University of Athens
agsp96@media.uoa.gr

Iraklis Varlamis
Athens University of Economics and Business
Department of Informatics
varlamis@aueb.gr

Abstract:

Conducting Web research has been notoriously difficult due to a number of interrelated reasons, most of which are inherent to the nature and architecture of the Web. As a consequence, the selection and gathering of appropriate samples for quantitative research, as well as the employment of qualitative methods at large, are confronted with severe obstacles, in most cases bypassed only by ad hoc solutions.

Feeds are XML-based technologies used mainly for syndication by means of sharing information in a structured format. As such, feeds have to be considered more as a communicative and content-sharing tool than anything else; their proliferation depends on the development of feed reading software such as aggregators, as well as the usage of the latest generation of feed-enabled browsers. The fact that feeds are part of the semantic web bunch of technologies indicates the appropriateness of feeds as methodological tools for focused web research.

In this paper the various aspects of the utilization of feeds as research tools are examined in order to provide suggestive solutions in terms of suitable and efficient tools to a segmentation of the methodological barriers posed to web research.

The argumentation expands from the current use of feeds as an information sharing technique to the proposed one as a methodological tool, leading to conclusions which favour the selective usage of feeds as tools for web research.

Résumé

Mener une recherche sur le Web a été extrêmement difficile pour différentes raisons étroitement liées, la plupart étant inhérentes à la nature et à l'architecture du Web. En conséquence, la sélection et la collecte d'échantillons appropriés pour une recherche quantitative, tout comme l'utilisation globale de méthodes qualitatives, sont confrontées à de sérieux obstacles qui, dans la plupart des cas ont été contournés grâce à des solutions ad hoc.

Les fils sont des techniques utilisées surtout pour les sites Web, blogs et syndication de nouvelles par le moyen d'un partage de l'information dans un format structuré. Comme tels, les fils doivent être essentiellement considérés comme un outil de communication et de partage de contenu. Le fait que les fils fassent partie de la grappe sémantique de technologies du Web, manifeste la justesse de l'utilisation de ces fils comme outils méthodologiques pour une recherche centrée sur le Web.

Dans cet article, les différents aspects de l'utilisation des fils comme outils de recherche sont examinés afin de fournir des solutions suggestives (bien qu'indicatives) sur le plan des outils adéquats pour une segmentation des barrières méthodologiques posées à la recherche Web.

L'argumentation se développe à partir de l' utilisation courante des fils comme technique de partage de l'information, jusqu'à celle proposée d'un outil méthodologique menant à des conclusions qui favorisent l'utilisation sélective des fils comme outil pour la recherche Web.

---

1. Introduction

The rapid development and the extended use of feeds in the contemporary Web environment provide enough evidence that XML-based technologies used for the delivery of content to end users are of stable foundation and researchers may rely on them in order to collect data for their research related tasks and activities in medium-dated time scale.

In parallel, Web analytics at large has been a controversial –yet, omnipresent– issue in the relevant discussions within the Web development community. Moreover, Web analytics is of great interest among the marketers (Scott, 2007), especially if one considers the importance of e-commerce in a world of globalized markets and the fact that information collected via the analysis of website usage statistics is the sole most decisive factor in the process of decision making as far as online marketing is concerned.

The above mentioned topics –proliferation of feeds and Web analytics– coexist as key parameters in the already present Web of services and social networks and constitute main gateways for the study of the Web as introduced in the early descriptions of the emerging Web Science discipline (Berners-Lee *et al.*, 2006). Following this particular line of thinking it is clear that feeds and Web analytics intersect, while the former may be an input resource for the analysis conducted by means of the latter in terms of methodology, technique or, in the simplest case, in purely instrumentalistic fashion (i.e. considering feeds as a data collection tool).

In this paper the current status of RSS feeds and Web analytics is briefly presented in order to examine the bidirectional relation of the two segments and in particular the possible ways in which feed readability may be measured (analytics-to-feeds direction) and the more obscure, yet more challenging, ways in which feeds may be exploited as tools for Web research (feeds-to-analytics direction).

The conclusions presented in the last part of this theoretical discussion favour the experimental use of feeds as input mechanisms for Web research and suggest that the various implications of such an approach should be further investigated, taking into consideration the optimistic estimations regarding the current status of feed diffusion and the predictions for the increasing importance of Web analytics. Though there are several unresolved issues concerning both the technical and the theoretical aspects of the proposed usage of feeds, there is adequate evidence supporting the core idea presented in this paper, namely that RSS feeds as tools for Web research can be exploited in a statistically sound way.

2. Web Analytics at Large

It was as early as 1995, just on the dawn of the Web explosion, that Dr. Stephen Turner introduced one of the very first software packages for log files analysis and reporting (Kaushik, 2007:3). Web analytics -as a specialized field of statistical analysis, as a medium and content consumption indicator and as a quite profitable profession- remains ever since those early days an integrated part of the evolution of the Web. Objectives set under cryptic acronyms such as KPI (Key Performance Indicators) (Clifton, 2008:154) have become the driving force behind both website development and marketing decisions.

Essentially Web analytics may be considered as the equivalent of traditional media metrics, in the sense that they provide information regarding the performance and popularity of websites. At the same time, Web analytics is an open field for Web and media research, since the key data gathered by the various data collection methods employed in analytics reveal significant trends and usage patterns in fields like content consumption, user behaviour and online habits.

From a technical perspective, Web analytics has evolved and keeps growing in order to meet the different challenges posed by modern Web 2.0 technologies and especially rich media content and Ajax-related content delivery in which the single building element/unit of the Web, namely the webpage, is no more present in the way it used to be up to even a few years ago. The most widespread methods for gathering data are the parsing of web server software log files, the client-side methods which are employed by the use of JavaScript tags, cookie-based methods, and lately, more sophisticated complementary methods which utilize a combination of the previous ones in the form of services rather than standalone applications. Those development stages are not separated from each other, but rather each stage has provided different tools and techniques which coexist as alternative approaches to the same ultimate goal, which remains the gathering of accurate and reliable data regarding the use of a website by its visitors.

All the interesting research work in the personalisation of content delivery and the generation of content recommendations for the users has been based on the analysis of the usage data of a web site (Eirinaki & Vazirgiannis, 2003). In the case of Web 2.0 applications existing solutions are not viable any more. Especially in the case of feed readers, users are able to access feeds on different servers and as a result their browsing history is fragmented into many web log files. In order to be able to process usage data in this case, we need access to all the servers that host the web log fragments. A more applicable solution is to use a mediator service. The service can either serve the original content and record every access to it (e.g. the personalized feed aggregator at www.feeds2.com), or it can simply collect and processes web usage data from the various web servers. In the latter case, servers act as beacons that transmit every transaction to the mediator (Laratro, 2008).

Though it can be safely claimed that Web analytics has reached a mature stage of development, yet a wide consensus on the definition of Web analytics has not yet been reached. The Web Analytics Association provides the following definition in a working paper which has been released lately for public comment: "Web Analytics is the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage" (Web Analytics Association, 2008).

Unlike traditional media metrics, Web analytics, due to the interactive nature of the medium, may be effectively –yet, indirectly– used for the purpose of obtaining useful research data besides the performance and usage of the medium-unit (i.e. the website) per se. An example would be the posting of comments in a particular topic, which in turn, in addition to information about the popularity of the article, provide data for content analysis and other qualitative research methods. Such assumptions lead us directly to the core problem online research has been confronted with, since day one of the Web: sampling methods and in particular the quest for a safe way to ensure that the collected data originate from a randomly selected portion of the population bearing its typical characteristics. In the example just given, commentators are not randomly selected, which means that the sample is biased and therefore no inductive statistical methods may be employed. The example provided confirms the view that "essentially, in any form of online testing, you have an extremely limited opportunity to sample your audience in a way that would constitute a truly valid scientific experiment" (Eisenberg *et al.*, 2008:248).

Web analytics differ from the measuring of traditional media in several respects, among which the following is fundamental: in traditional media the popularity of the medium is measured using data collected only from a fraction –usually a small sample– of the population of users. Even in the case of print media, where one may consider as a strong indicator of popularity the number of subscriptions or the number of copies sold, the readability of the medium, defined as the number of readers who actually read the medium, cannot be accurately measured and it should not be considered as being equal to the number of copies delivered (i.e. all the members of one family read the same copy of a newspaper). In other cases, as in radio ratings, for example, the measurement takes place in the form of short interviews or through questionnaires filled in by a rather limited sample of the population. In all cases the results are being generalized to the whole population by inductive methods, under the non-explicit presumption that the sample was indeed random and selected through scientifically acceptable methods (which is not always the case).

The following extended excerpt from *Actionable Web Analytics* by Burby and Atchison (2007:13) gives us a vivid summary of the situation that existed in traditional media and makes reasonable predictions for the forthcoming media landscape:

> "You may recall how TV ratings were conducted many years ago: Families received a diary in the mail in which they were to record the shows they watched each evening (the *Nielsen family*); they mailed the diary back every few weeks. From that highly scientific and precise survey the TV networks determined the ad rates for their shows. Noone questioned the model because it was the best the medium offered. Today we have digital cable and Internet Protocol TV (IPTV) bringing all the media channels closer together in both spirit and technology. The lessons learned in measuring the Internet over the past decade will provide the framework for eventually measuring all media".

On the contrary, the data used in Web analytics are directly collected from each and every visitor of the website reflecting accurate metrics in absolute units as opposed to the averages and percentages in which the results of the traditional methods are expressed. The

fact that the data in the case of the Web are collected from the whole population of visitors increases significantly the reliability of the results, at least when compared to the results obtained by such methods as the filling in of questionnaires. However, on the other hand, we have little to none information regarding the demographic characteristics of the particular population, which in turn means that the sampling problem remains unsolved. In fact we do have rich, detailed data, but we cannot use them for research other than for the popularity and effectiveness of the particular website from which they originated.

3. Feeds Proliferation

The basic idea behind news feeds is clear and responds to a real communicative need on the side of the end-user: there should be an easy way for developers to deliver and for users to get frequently updated content from web resources other than visiting the site and trying to figure out which pages have been updated and if any new content has been added. In order to reach this goal, various tools have been used dating back as early as the first years of the Web. Even mailing-lists, newsletters and e-mail alerts have been used for the same purpose with various rates of success, the newsletter being for many years the most effective solution as far as the end-user is concerned. On the side of Web developers a number of scripts, usually in the form of cgi programs, written in Perl, which at the time was the dominant server-side scripting language, promised to fetch news and headlines from major news websites and portals. Nevertheless it was only the advent of XML-based formats like RSS and, more recently, Atom, which made both ends meet and provided developers with a solution that is easy to integrate to their websites and end-users with the latest news and updates of their choice right on to their desktops. One of the most basic promises of the Semantic Web, namely machine-readable content being automatically delivered to end-users, was becoming a reality (World Wide Web Consortium, 2001). According to Hammersley (2005:1) "the original, and still the most common, use of RSS and Atom is to provide a content syndication feed: a consistent, machine-readable file that allows web sites to share their content with other applications in a standard way. Originally […] this was used to share data among web sites, but now it's most commonly used between a site and a desktop application called a reader."

The issue of Web standards, as a means to reach a common set of rules for content development and services on the Web (Zeldman, 2003), has been widely discussed and remains a critical one (Giannakoulopoulos, 2007). The same holds true in the case of feeds in which no common standard has yet evolved, though there have been serious attempts towards this direction (Internet Engineering Task Force, 2005; Harvard Law, 2007). Up to 2006 there were nine different versions of RSS only (i.e. excluding Atom), most of which were considered obsolete, though still in use (Johnson, 2006:68-69). "From a developer's perspective, the RSS situation looks like a nightmare, but it's really not that bad. The good news is that if you stick to the basic elements -<item>, <title>, <description, <pubDate>, and <link>- or you use a good parsing library, you will be able to parse RSS with relative ease." (ibid., 69-70). The documented relative easiness to create and parse feeds is one of the key points in the proposed use of feeds as tools for data collection.

The proliferation of feeds in contemporary Web is for the most part related to the expansion of the blogosphere, since it was the community of bloggers which wholeheartedly embraced content syndication in order to strengthen the bonds between its members. In addition, from a purely technological perspective, the direct relation between blogs and feeds originates from the fact that most well-known and widely used blogging

software packages provide feed mechanisms by default and bloggers get accustomed to using them by user-friendly administration interfaces.

The technical characteristics of weblogs determine, more or less, the availability of communicative options bloggers may exploit. On the one hand, the dynamic features of the interface provided to the end user –i.e. the blogger herself– facilitate daily, or even more often, postings to a previously unimaginable degree, while, on the other hand, syndication practices, fundamentally abutted against the respective technologies, support extensive diffusion of content and opinions. The following brief theoretical analysis of the technical features of blogs and feeds in particular, takes advantage of the conclusions drawn by Killoran and attempts to parallelize in a causative manner the rapid spreading of weblogs with the advancement of dynamic web technologies during the last few years of the previous decade.

Killoran, referring to personal homepages and weblogs as different genres of personal websites, mentions that "over the past decade, two different genres of personal Web publishing have experienced different histories and offer different promise for popular participation" (Killoran, 2003) and concludes that "in the new media environment favoring currency and enabling easy contact among publishers, it is, of these two genres, the blog that can better sustain a more distinctive, more socially responsive means of civic engagement." (ibid.). Speculating on both Killoran's conclusion and the technical features of weblogs, we suggest that –at least partially– the above mentioned viability and precedence of weblogs is the result of the dynamic web development techniques employed in weblog software and applications.

Following the arrow of time, personal homepages may be thought of as some kind of historical predecessor of weblogs. The use of websites for publishing personal information dates back to Web's infancy and the first form of self-exposure to the Web came in the shape of personal homepages. The vast majority of personal homepages were constructed in plain HTML, i.e. a static markup language enriched with hyperlink functionality. HTML's simplicity has been praised as the foremost important factor for the rapid expanding of the Web, since the ability to construct hypertexts in a meaningful manner followed the fully descriptive tags of HTML. From a communicative perspective, HTML succeeded as a global authoring tool by offering the opportunity to virtually anyone with a personal computer and an Internet connection to effectively make her appearance in the electronic public sphere. Although hypertext, as deployed in the early static web constructions, was a major step forward compared to previous linear narrative models, the evolution towards more dynamic solutions proved that pure static constructions could not fully exploit the creative potential underlying the concept of hypertext. Personal homepages lacked regular updating and had little to none new content added after the initial launching, the reason being that static constructions require laborious HTML coding for the formation of functional hypertext and a whole different process of uploading files to the web server –in most cases through FTP– in order to become publicly available. Weblog software and applications are predominantly based on dynamic technologies. This fact makes both authoring and publishing a lot easier for the laymen and removes the formatting load from the author to the automation software.

An interesting feature of blogs, which is tightly coupled with feeds, is the use of tags. Tags are usually single words or word pairs, which are used to annotate the content. On top of tag information have been designed many knowledge management solutions and have been developed many interesting applications. With the use of tags, bloggers are able to give the key terms of their posts, but also to annotate the work of others thus creating a social bookmarking repository (Heymann *et al.*, 2008). Tags and the original content can

then be organized and presented in various ways (i.e. clouds of tags, graphs and taxonomies called folksonomies etc.). Tags can also be used as a replacement for the content, especially in information retrieval tasks and when the processing of the original content becomes very laborious, or when the original content is not in text format (i.e. photos).

Several applications have been developed that use the tag information to filter or re-organize feed contents. An example is "RSS feeds for tags" service of Amazon (Amazon, 2008), which automatically generates feeds for keywords in a web page. The feeds contain articles tagged with the specific keywords.

As mentioned above, blogging software automatically publishes an XML-formatted feed, which may be used to keep users updated without having to revisit the website, and simultaneously provides researchers with valuable and, most importantly, well-organized machine-readable data ready to be used in statistical packages. Though further elaboration is obviously needed in order to reach scientifically acceptable solutions, it is clear even to this point that the proliferation of feeds offers a wide range of potential data worth exploiting for research objectives.

4. Exploiting Feed Data in a Bidirectional Manner

Web 2.0, which relies on the XML family of technologies –including, of course, feeds–, designates a major challenge for the statistical analysis of the kind Web analytics performs. Web 2.0 is a phrase attributed to Tim O'Reilly and in effect, according to Clifton (2008:269), "is a buzzword for the next generation of browser applications. […] The irony is that the technology that drives Web 2.0 is part of today's Web 1.0 technology and has been around for many years –that is, JavaScript and XML. As such, Web 2.0 does not refer to any technical advancements of the Web or the Internet infrastructure it runs on, but to changes in the way the medium is used". Therefore, since Web analytics measure primarily the usage of websites it is pretty obvious that changes to "the way the medium is used" cause analogous changes to the measurement methods.

There are two main axes along which the measurement of services in the contemporary Web differs from that of previous years: the first has to do with Ajax-like technologies in which only part of the page is updated without fulfilling the whole request-respond cycle. "The techniques employed when developing a website using Web 2.0 technologies separate the components of data, format, style, and function […] This has tremendous implications when it comes to defining KPIs, as the concept of page view all but disappears" (ibid., 270); the second, which is directly related to the issue under discussion has to do with the fact that due to advanced syndication mechanisms large portions of content are not consumed within the context of the website in which they were originally published but in other locations on the Web (which syndicate to the content providing site) or in desktop applications. Therefore, for the scope of our proposal we will have to approach the cross-section of Web analytics and feeds in a bidirectional manner, in an attempt to identify methods to measure feed popularity and readability (step one) and to find ways in which researchers may exploit feeds in order to collect data for their research efforts (step two).

As far as the measuring of the feeds is concerned, "it is surprising that no vendors have stepped up to provide a solution to measure content syndicated via really simple syndication (RSS)" (Peterson, 2005:191). Since Peterson's aphorism in 2005, there have been efforts to measure indeed the readability and success of feeds, yet no stable solution has been found, most probably due to the lack of a common standard format for news

feeds. As Kaushik (2008) puts it in his Occam's Razor blog "the challenge on a blog is that the content is on the website and also available via RSS. So how do you know true content consumption?" The answer to that question is not straightforward; we will have to blend various different options in order to reach a more or less accurate estimation, rather than absolute raw results.

Peterson (2005:45) suggests that blog owners and feed providers could build their own tracking application and provides sample code in JavaScript and Perl; on the other hand Kaushik (2007:322-323) considers adequate the results provided by vendors and services such as Visual Sciences and FeedBurner. Due to the constraints mentioned above, along other more technical ones (e.g. the fact the standalone feed readers do not accept cookies) it is not feasible in the case of feeds to have deep analysis and detailed reports of the kind that can be generated for websites. Still, either of the methods suggested in the bibliography would provide the feed owner with all of the core metrics including number of subscribers and visited links.

The single most beneficial element behind the idea of using feeds as data collectors is that the whole process is to a large extent automated. Most latest generation statistical packages, spreadsheets and RDBM systems support direct input in XML compatible formats which in turn means that the whole burden of data entry, as well as a considerable part of the effort needed for data collection may be transferred from the researcher to software applications. Still the most difficult and challenging part is not only to get rid of some laborious extra work but to exploit feeds as methodological tools which are integrated to the initial planning and design of online research.

Let us provide an indicative example of such an approach. Suppose that a team of social researchers specialized in gender studies would like to compare the number of female and male staff members appointed to Greek Universities over a given time period. Since appointments to the position of teaching staff member are official documents accessible to the public, the team of researchers would have to contact all the Greek Universities and get the data needed to answer their research question. By the feeds approach, under certain preconditions, the most essential of which is that Universities would have to provide feeds of their official announcements, the whole process would be much easier and the data needed would be readily available for statistical elaboration.

An analogous example may be constructed the other way around: suppose that the team of researchers has subscribed to a number of reliable news sites. Then, by conducting content analysis or even plain keyword search over the title element of the feeds, the team may obtain data regarding, e.g., the number of times the name of a candidate appears in the news headlines over an election campaign period in every medium of the sample and draw useful conclusions from the comparison.

In another variation, the research team may construct feeds for its own use exclusively built for research purposes. For example a feed providing data about search results could keep track of specific search topics, which in turn may be quite useful as input data for researchers working on the topic which has been tracked. Hammersley (2005: 196-197) provides a fully-coded example on how to implement a search results feed using Google's API and SOAP envelopes to deliver data from search results. In his own words "whether you are trying to find things out or monitoring for other people's discoveries, an RSS feed is the perfect way to do it".

Obviously using feeds as tools for data collection is not appropriate for all kinds of online research. In fact there are certain research areas in which feeds as tools for data collection are quite acceptable whereas in other areas they may cause more problems than the ones the method is trying to solve. In cases where direct input from the research

participants is needed, as in the characteristic example of online questionnaires, feeds cannot be of any considerable help, whereas in research conducted to published data or archives of related documents, aiming to measure popularity or to gather data for content and text analysis, researchers may benefit a lot from the proposed feed technique.

5. Conclusions

The conclusions to be drawn from the above theoretical discussion favour the use of feeds as tools for web research. At the same time it is evident that there are a number of issues which require further discussion in order for this proposal to be effective and lead to statistically acceptable results, regardless of the techniques which may provide ad hoc solutions to particular problems.

The sheer fact is that Hammersley (2005) describes blogging as extremely popular due to the fact that "at time of writing, in late 2004, there are an estimated four million weblogs being written worldwide", while that number sky rocked to 70-million-plus blogs according to Kaushik (2007) in a less than a 3 years period. The vast majority of these blogs offer RSS or Atom feeds by default, as the relevant mechanism is inherently embodied to the kernel of the blog code. All these feeds constitute potentially a research corpus of well-organized data readable by machines that may be further organized in subcategories according to the needs of each particular research objective. Whether researchers prefer -from a purely methodological point of view- to work with random samples or to define a whole population in terms of its specific characteristics, the information provided by feeds may be considered a valuable input resource for data collection.

Furthermore, this approach adds extra value to the economic efficiency of the research, because it eliminates data entry costs and other time and effort consuming activities, since XML formatted data, such as feeds, are machine readable and well supported by most statistical packages, spreadsheets and database management systems. Therefore researchers may concentrate on the key issues and questions of their research, neglecting annoying technicalities such as data format conversions and incompatibilities among the various applications used for tasks related to data manipulation and statistical elaboration.

Obviously, all these benefits come at a cost: the standards for syndication technologies should not be considered stable and this in turn weakens the possibilities for long-term investigatory engagements and meta-research, which is not the case in standard data collection techniques; as Kaushik (ibid:322) claims "RSS is still not very integrated into most websites and the analytics tools that are being used by the websites. There is no standard when it comes to tracking RSS". The lack of standards also restricts the comparability options among different sets of research results. By evaluating the significance of the drawbacks of the proposed solution we may conclude that standardization of the content syndication formats –which may be reasonably anticipated in the foreseeable future– will remove a considerable part of the implementation deficiencies and make the use of feeds as tools for data collection more reliable and accurate.

The epigrammatic conclusion of the discussion supports the use of feeds as tools for web-based research under certain preconditions directly related to the type of research and the kind of information needed for each research effort; it also proposes further exploration of the issue, both in theoretical and actual implementation levels, by means of paradigmatic cases.

Bibliographic References

1. Amazon, *Feeds for tags service*. Available from: <http://www.amazon.com/gp/tagging/rss-help.html > [Accessed October, 1st 2008].

2. Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N. & Weitzner D. *A framework for web science*. Hanover, MA: Now publishers, 2006.

3. Burby, J. & Atchison, S., *Actionable web analytics: using data to make smart business decisions.* Canada: Wiley Publishing, 2007.

4. Clifton, B. *Advanced web metrics with Google Analytics.* Canada: Wiley Publishing, 2008.

5. Eirinaki, M. and Vazirgiannis, M., *Web mining for web personalization*. ACM Trans. Interet Technol. 3, 1, 1-27, 2003

6. Eisenberg, E., Davis, L. & Quarto – von Tivadar, J. *Always be testing: The complete guide to Google website optimizer.* Canada: Wiley Publishing, 2008.

7. Giannakoulopoulos, A.P. 'Web standards and consensus: ensuring the universality of Internet applications', *Communication Issues*, 5, 2007. pp. 33-45.

8. Internet Engineering Task Force (IETF) *The Atom Syndication Format* [Internet]. Available from: <http://www.ietf.org/rfc/rfc4287.txt/> [Accessed August, 26th 2008], 2005.

9. Johnson, D. *RSS and atom in action: Web 2.0 building blocks*. Greenwich: Manning Publications, 2006.

10. Hammersley, B. *Developing feeds with RSS and Atom*. Canada: O' Reilly Media, 2005.

11. Heymann, P., Koutrika, G., Garcia-Molina, H. *Can Social Bookmarking Improve Web Search?*. First ACM International Conference on Web Search and Data Mining. 2008.

12. Kaushik, A Occam's Razor [Internet]. Available from: <http://www.kaushik.net/avinash//> [Accessed 28 August 2008], 2008.

13. Kaushik, A. *Web analytics: an hour a day*. Canada: Wiley Publishing, 2007.

14. Killoran, J. B. "Homepages, Blogs, and the Chronotopic Dimensions of Personal Civic (Dis-) Engagement." *Rhetorical Democracy: Discursive Practices of Civic Engagement*. Ed. Gerald A. Hauser and Amy Grim. Mahwah, NJ: Lawrence Erlbaum Associates, 2003.

15. Laratro, J., *Google Analytics - A Practical Guide to Implementation*. White paper by MoreVisibility Available from: < http://www.morevisibility.com/whitepaper/GoogleAnalytics.pdf> [Accessed October, 1st 2008].

16. Peterson E.T. *Web site measurement hacks: tips and tools to help optimize your online business.* Canada: O' Reilly Media, 2005.

17. Scott Meerman, D. *The new rules of marketing and PR: How to use news releases, blogs, podcasting, viral marketing and online media to reach buyers directly.* Canada: Wiley Publishing, 2007.

18. RSS 2.0 at Harvard Law *RSS 2.0 specification* [Internet]. Available from: <http://cyber.law.harvard.edu/rss/> [Accessed August, 28th 2008], 2007.

19. Web Analytics Association (WAA) *Best Practices* [Internet]. Available from: <http://www.webanalyticsassociation.org/articles/> [Accessed August, 26th 2008], 2007.

20. World Wide Web Consortium *Semantic Web Activity* [Internet]. Available from: <http://www.w3.org/2001/sw/> [Accessed September, 3rd 2008], 2001.

21. Zeldman, J. *Designing with web standards*. New Riders Publishing, Indianapolis, 2003.