

# **Digital Literacy: Web search ecology and some surprising conclusions about finding and promoting educational resources on the internet**

Laurence Cuffe<sup>1</sup>  
*National College of Ireland*

*(Received March 2014; final version received September 2014)*

## **1. Introduction**

An ecology is a system of interacting and or competing organisms or entities. Alternatively, it may also describe the study of such interactions or competition. In this paper I loosely consider the system of educationalists searching for content to use in their teaching practice to form such a system.

Having returned from a course on ICT for social and group learning delivered by the company Smart Solutions (Vassallo 2010) in Malta, I decided to use what I had learnt to build a number of websites publicising the computer tools I had learned about. In choosing free websites, I was choosing platforms which would be available at no cost to educators, as I was hoping to promote teachers' active participation in the online community, which had been a theme of the course. Using free internet sites, I built a wiki (Cuffe 2011a), on Wikispaces (Anon 2010b), a blog (Cuffe 2010a) on the WordPress.com site (Anon 2011g), a Google group (Cuffe 2010c), on Google groups (Anon 2011c), and a Trailmeme (Cuffe 2011c), on the Trailmeme site (Anon 2011e). I also built a second blog (Cuffe 2010b) on Blogger (Anon 2011a) and shot a number of Videos (Cuffe 2010d) which I placed on the You Tube site (Anon 2011i). Finally, I built a Prezi (Cuffe 2010e), which I placed on the Prezi site.

Some of the sites I created initially placed highly in Google search however web traffic did not seem to correspond in any clear way to my sites rank in search. This was surprising, however I noted that traffic would peak whenever I mentioned a specific site on an online forum, or on an email chat group, and this peak could be many times larger than the regular level of traffic coming in, presumably, from general search (see Figure 1). The rest of this paper describes what I did to explore this further.

## **2. Methodology**

Initially I looked at the usage statistics, where they were available, for a number of web resources I built. In examining the traffic data for these sites with a view to estimating their relative effectiveness, I noted that in web search both Prezi and Trailmeme sites ranked highly, even when there appeared to be little traffic passing through them. Stand-alone blog entries did not fare well.

A more detailed examination of traffic logs revealed that promoting the site via other channels such as mail rings and focused email groups such as the CESI List (Computer Education Society of Ireland

---

<sup>1</sup> Corresponding author. Email: [cuffe@mac.com](mailto:cuffe@mac.com)

2011) resulted in a sharp spike in user traffic which then trailed off over the next day or two (Figure 1).

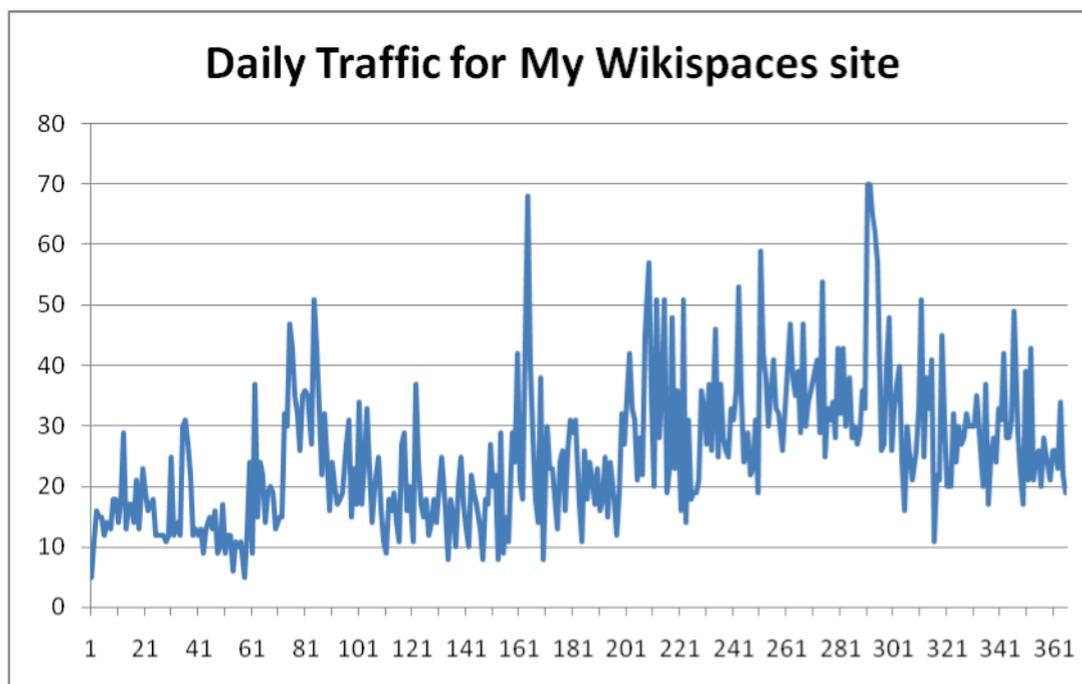


Figure 1: Unique visitor numbers for my Wiki spaces website for 2011, showing the peaked nature of the distribution. The peak around day 165 corresponds to a promotion of the website via the computers in education society of Ireland mailing list other peaks correspond to similar mentions on other media.

Other strategies such as promoting the resource via Twitter (Anon 2011f), and in comments on popular educational blogs were less effective for me, though the effectiveness of such methods would depend very much on the prominence of a user’s digital identity. As I currently have only 227 Twitter followers, the Twitter result is unsurprising, though Twitter might be more effective for Lady Gaga, who currently has 24,000,000 followers. This conclusion also held for the use of bulk email. A mail shot to a select group of 70 educationalists involved in Higher education resulted in no detectable increase in site traffic. The use of Facebook, Pinterest, or Google+ also has potential for users with a sufficient number of followers on these platforms.

As these peaks could be many times larger than the traffic which was arriving at my educational sites via generalized search, I then set out to determine if this behaviour was normal for other educational web sites, or was anomalous. To perform this analysis I looked for other educational web sites for which traffic statistics were freely available. Because Wikispaces gave the most comprehensive breakdown of web traffic and was, as far as I could tell, the most popular platform I used to disseminate this information (Table 1), I concentrated my analysis on looking at the traffic statistics available publicly for a sample of Wikis built on the Wikispaces site (Anon 2010b).

2011 Cumulative figures		
Site	Total Visitors	% of total
Wiki	9325	77.31%
Trailmeme	1400	11.61%
Blog2(wordpress)	537	4.45%
YouTube	557	4.62%
Prezi	177	1.47%
Blog1(blogger)	64	0.53%
Google Group	2	0.02%
Total	12062	100.00%

*Table 1: A comparison of traffic flows to a selection of websites built to disseminate information about tools for social and group learning.*

This phase of the traffic analysis, looking at daily statistics for 43 educational wikis for 2011 covered 2.7 million visits to this collection of sites, by so called “unique visitors”. I used this data to establish that arrival statistics were significantly more peaked than one might expect if visitors were arriving at random, and were thus more likely to result from specific mentions or promotions of the resources, as opposed to resulting from generalized random search.

After completing this analysis, it occurred to me that perhaps by looking at just educational Wiki sites, I was being too focused in my research. I then used Needlebase, which allows you to automate data collection from a set of websites to collect data from 28 other educational web sites over a 45 day period. These sites included educational videos, university websites, online collections of resources and educational blogs. This data set covered about 1.6 million site visits, and supported my previous conclusions.

## **2.1 Details of data analysis**

Educational wikis visitor statistics were gathered from a purposive selection of the online list of educational Wiki’s created by Steve Hagedorn, (Hargadon 2011) using the following procedure.

Wiki descriptions on this site were initially analysed to determine if the primary role of the wiki was to disseminate information about web based educational resources to educators, if wikis matched this description, I then downloaded the unique visitor daily visits numbers for the year 2011 and transferred this data to Excel for further analysis.

The description was analysed before usage statistics were viewed or downloaded, so as to avoid any data driven bias in deciding which wikis to include in the data collection phase.

# **1. Statistical expectations of web site traffic**

Visual inspection of website traffic graphs showed sharp peaks superimposed on a baseline of random activity as shown in Figure 1. If traffic arrived at a website purely as a result of random searches I would expect such traffic to show a Poisson distribution. This distribution represents a good model for human activity where individual decisions to act are not related to others’ immediate actions. For the Poisson distribution the mean and the variance of the traffic data should be equal. This provides a simple test for the presence of such a distribution.

The data collected shows that traffic to the educational web sites under consideration was highly episodic, and came in very peaked flows. In an effort to characterise this statistically, I computed the mean and the variance of each sites visitor figures. This revealed that the variance of the traffic was at times orders of magnitude larger than the mean. I have taken the ratio of these two figures for each of the wiki spaces’ data sets, and plotted the log of this in Figure 2.

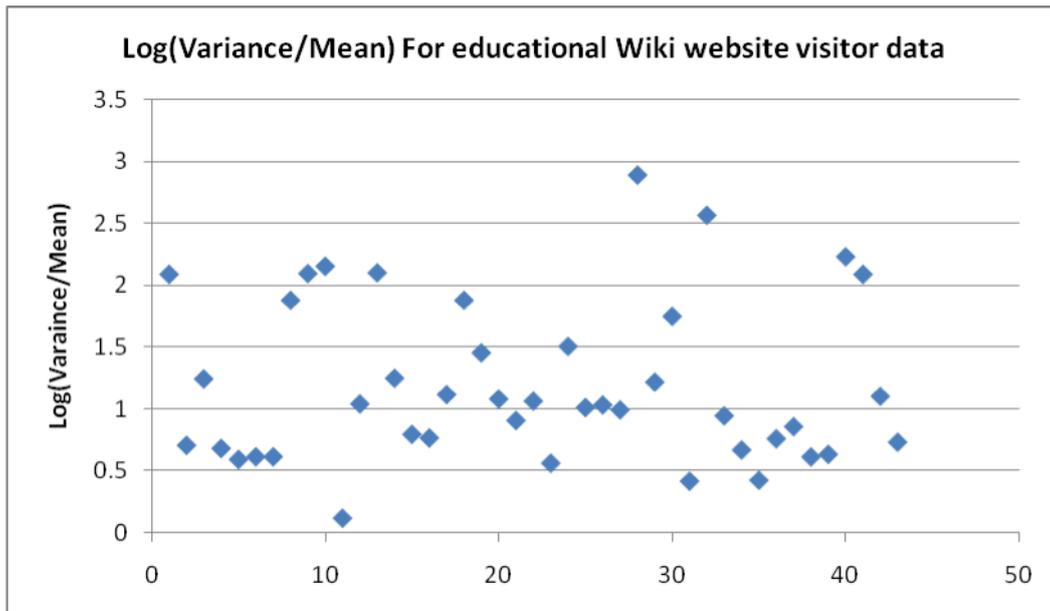


Figure 2: Plot of  $\text{Log}(\text{Variance}/\text{Mean})$  for daily incoming unique visitor numbers for 43 educational wikis. Note: If traffic arriving at these wikis was Poisson in character, the ratio of these two statistics would be one and the points would plot on the X axis, as  $\text{Log}(1) = 0$ . In all the data sets for which this ratio is plotted above, the variance exceeds the mean, and the distribution shows more variability than would be expected if the distribution was Poisson. This plot represents the activities of 2.7 million site visitors.

This plot is consistent with the hypothesis that traffic was being driven by specific promotions of the web site contents leading to very pronounced peaks in traffic as opposed to being a function of their position within search results.

Data for the other educational websites examined are plotted in Figure 3, below.

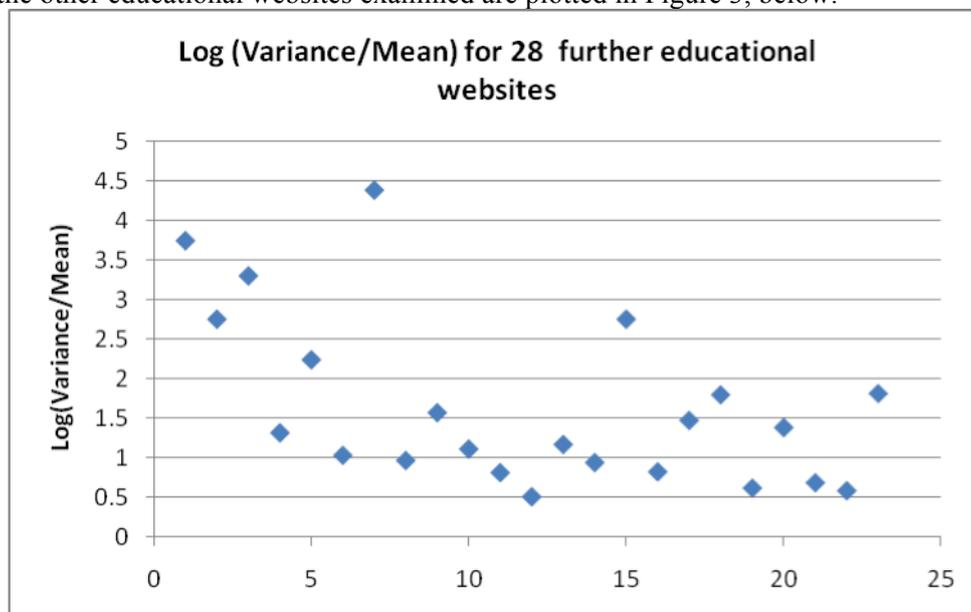


Figure 3: Plot of  $\text{Log}(\text{Variance}/\text{Mean})$  for daily incoming unique visitor numbers for a selection of 28 educational (non-wiki) websites. Note: If traffic arriving at these sites was Poisson in character, the ratio of these two statistics would be one and the points would plot on the X axis, as  $\text{Log}(1) = 0$ . In all the data sets for which this ratio is plotted above, the variance exceeds the mean, and the distribution shows more variability than would be expected if the distribution was Poisson. This graph represents the activities of 1.6 million site visitors.

Other literature supports this conclusion. A paper (Kumar & Tomkins 2009) which analysed 50 million browser page views concluded that only 20% resulted either directly or indirectly from search engine results. Here an indirect arrival refers to a user who may click on a search result and follow the link deeper into the web to arrive at the final web page. In Kumar and Tomkins paper data was gathered from logs of all internet activity of users who installed web-tracking software in their browsers. The high variability of educational website traffic is also supported by earlier work (Meiss et al. 2005). Meiss et al. analysed large scale (740 million data transfers) traffic flows on the academic Internet2 network, where they found that

For such a distribution, the second moment:

$$\langle n_s^2 \rangle = \int n_s^2 P(n_s) dn$$

eventually diverges; the standard deviation is not an intrinsic value of the distribution and is only bounded by the size of the statistical sample. (Meiss et al. 2005, p.513)

This sample is interesting because it is very large, and consisted solely of educational traffic. The website traffic I analysed covered a year for the wiki data, and 45 days for the other educational website data. This data could well be describable over the long term by the same sort of scale free behaviour as that described in the quotation from Meiss et al, given above.

This analysis supports the conclusion that educational website traffic is not, at this point in time, derived primarily from search engines. This conclusion could have been drawn from website analytics, looking at the referring site information for arriving visitors, however the conclusion reached above is more robust, insofar as it would capture data for visitors arriving via referral chains, referred to above, as well as direct arrivals.

## **1. Survey**

In addition to this statistical analysis, nine short structured interviews of educators were carried out. Data from these interviews was used to triangulate the information gathered in the previous phases of the research.

A purposive sample was devised (Silverman 2000) sampling a cross section of educators and trying to avoid the bias which could arise by either enrolling educators via online forums, or those who were active in an e-Learning context. The sample was chosen to provide a broad cross section of educators using the web to teach today, and consisted of individuals drawn across a range of educational contexts. With this in mind one primary teacher, two secondary teachers three third level lecturers, and three adult education teachers were chosen. All interviews were conducted confidentially, and the names of interviewees are withheld by mutual agreement.

This group were asked about their Internet use in teaching and were then queried as to how they located resources which they used in class. In asking specifically about resources used in class, I wished to exclude search methods and methodologies used in educational research as distinct from teaching practice.

### **4.1 Results**

Only one respondent indicated that Google was their primary method for locating resources. When asked whether they would locate video material via Google, this respondent said that they would go directly to you tube. This was the only respondent for whom generalized search was the primary source of teaching material. All other respondents indicated that they located resources via personal recommendations or “via a web site”.

As this research progressed I became fascinated by the low ranking that Google had in educators’ research strategy and I asked more specific questions as to why it was not used. Responses varied, but

are perhaps best summed up by one respondent who said “It’s all a load of rubbish; you get nothing but rubbish sites when you use it”. This, I think, captures a general sense of frustration with the large number of not very useful sites which you get if you put in search terms such as “Mathematics” and “Educational sites”. While these data support my conclusions, one very informal additional piece of research provides a small counter note, but also increased my understanding of why this result seemed surprising. At a conference for adult educators I asked a group of the participants about their strategies for locating educational material. An informal show of hands in a workshop with 18 educators present indicated that about half of the participants identified with Google or another search engine as being their primary source of educational material. In explaining this divergent result, I note that this group were active researchers attending an international conference. I would expect this group to be more skilled in choosing appropriate search words, and using search engines effectively to gather information.

No respondent mentioned more subject specific web portals, etc.; however I am certain that a number of them would use more specialized web portals for identifying research literature as part of their academic research.

## 2. Search engine optimization

Existing received wisdom (Fetterly et al. 2004), (Stanford School of Medicine 2011), (Wilson 2011) suggests that the most effective way to maximise web site traffic is to maximise the sites exposure via search engines such as Google and Bing. There is a lot of research and advice available (King 2008), (Langville & Meyer 2006),(Evans 2007) on how to improve your page rank with Google, and Google themselves also have an excellent publication on this topic, the “Google Search Engine Optimization Starter Guide” from Google (Anon n.d.).

In my initial search results both the Prezi and Trilmeme sites ranked very highly, while the Wiki spaces and Blog based sites did worse. From this I would conclude that using Prezi or Trilmeme to present an educational idea or resource will give higher initial visibility in Google search results than presenting your ideas via a blog post. Wiki spaces initially did not rank as well as did either of the two sites mentioned above, however as increasing traffic moved through the Wiki spaces site and as other authoritative educational sites linked to the Wiki spaces site it rose in the rankings. Parenthetically one of my wikis (Cuffe 2011b) has shown up consistently at or near the top of Irish Google’s searches for Irish educational resources, but has received little search directed traffic. A year later neither the Prezi nor the Trilmeme site rank as well which suggests that using the new hot platforms such as Pinterest, or “Scoop it” might receive the same novelty value boost as did the aforementioned sites, which were located on then quite novel and “trending” platforms.

I did not feel that high search rank was a key element in attracting site visitors, and, as this is a constantly changing domain of activity, I refer the reader to the previous publications and will not comment further.

## 3. Discussion

It seems that Google or Bing with farms full of servers surveying millions of websites should dominate any effective search for web resources. That this is not so is supported in part by the statistical analysis, and more strongly by the interview results.

I revert to the ecology analogy introduced at the start of this paper, and a well-known result from ecosystem modelling, Fisher’s fundamental theorem (Fisher 1930, p.37), (Price 1972). It states that “The rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time”, and this can shed some light on my conclusion. In a biological context, this theorem describes how a species in which succeeding generations have high variability, is more likely to

produce a particularly successful variant, which will come to dominate succeeding generations. I now apply this to web search.

If we consider the ecology under consideration to be the pool of educationalists searching for resources on the internet, and we partition this pool into two segments, those users using general search, and those using other means, Fisher's theorem implies that those using other means will have an advantage over those using general search, because of the greater range in the effectiveness of search using other means.

This increased effectiveness often takes the form of a symbiotic relationship between us, as educators, and our data sources, as we fine tune our data sources to favour those which are consistently useful, while discarding those sources which waste our time.

The General search landscape also changes. If similar work had been carried out five years ago YouTube (Anon 2011i) would have been a relatively unknown start up and AltaVista, (Anon 2011h) would have played a much more prominent role.

Moving on to the survey results I found in the interviews that trust and the ratio of high to low quality seemed to be an important factor in educators' choice of search strategies. Arrival statistics are also governed by Zipf's principle of least effort (Zipf 1949) which is significant in library science. Zipf's principle states that users are looking for the maximum amount of useful information in return for the least effort. This early statement about information seekers' behaviour has much in common with later models based on a foraging analogy (Chi et al. 2007). Following a link supplied by a known good source is in general more rewarding than using a generalized search engine such as Google.

It must be kept in mind that the survey referred to a very specific search exercise, looking for material which can be used in class. This is a distinct task from the academic literature search, or product search with a view to purchase, about which there is an extensive literature.

When I looked at website traffic data I sought to identify the factors which drove traffic to the websites I built on tools for social and group education. High search engine visibility was not a major factor. This conclusion comes from the observation that even when search engine rank was high, traffic remained low. An observation that peaks in website traffic correlated well with mentions in online communities of practice (Wenger 1998, 2006) such as Nings (Anon 2011d), and Mailrings such as Teachers Net (Anon 2010a), was explored by looking at daily unique visitor statistics for a selection of 43 other Wiki websites whose purpose was the dissemination of educational information and resources, as well as a 45 day sample of traffic to a diverse selection of 27 other educational websites. A comparison of these statistics with search traffic statistics for 40 educational terms gathered from a Google search statistics site, Google trends (Anon n.d.), supported the conclusion that traffic and search are not strongly correlated for these educational sites. Traffic on the educational websites was much more sporadic and showed a much higher variability than was seen in the statistics for search traffic where educational terms were used as key words.

Having arrived at this conclusion, I then asked why this should be so. Here the data is not as robust, being based on an examination of traffic logs for my own web sites and a series of semi structured interviews with nine other educators. The traffic data for my website indicated that traffic peaked in response to mentions of these websites either in online discussions or via Twitter. The interviews showed that users found generalized search engines such as Google frustrating due to the large number of irrelevant results returned. For example, a particular issue within the Irish context was the large number of results which had a poor fit to the Irish curriculum. Many results were described as being "American" in focus. This regional issue is probably representative of a larger problem, which is the difficulty of encoding in the search specific search target characteristics, such as appropriateness for a given curriculum.

One exception worth noting to the bias against generalised search was You Tube. All respondents indicated that if they were looking for a video clip to use in their teaching they would use You Tube to find it.

In the interviews respondents indicated that as an alternative to general search they found resources through recommendations from friends, or while on training courses.

The interview data is consistent with my previous conclusions, and indicates that practitioners prefer to use sites which have been recommended to them by people or sites which they trust, either face to face, online, or while on courses. Trust is a significant issue, and has been discussed in the literature (Sharratt & Usoro 2003) and more recently by the same authors (Usoro et al. 2007) who examine the factors contributing to effective knowledge sharing in online communities of practice.

My interview sample was small, and was chosen specifically to try and capture practitioners who did not use the internet a great deal. Results for a more web-proficient group of users would probably differ from those for the sample given above, as I would expect more proficient users to be able to use general search tools more effectively, and thus to use them more.

## 7. Conclusions

The traditional Community of Practice (Wenger 1998, 2006) based around a mailing list is still alive and well. While Google is often used to find short pieces of online material such as java applets, short video footage, or copies of standards, larger tools such as wiki software or websites with collections of educational resources are more likely to be discovered via a recommendation by a colleague, mention at a conference, or via online mailing lists or specialist websites such as Facebook (Anon 2011b). Trust and the ability of friends and acquaintances to filter the torrent of web content effectively seemed to be the factors driving this choice. This conclusion is supported by interviews with nine educators, and statistical analysis of educational web search and site usage statistics.

## References

- Anon, 2011a. Blogger. Available at: <http://www.blogger.com> [Accessed June 30, 2011].
- Anon, 2011b. Facebook. Welcome to Facebook. Available at: <http://www.facebook.com/> [Accessed June 30, 2011].
- Anon, 2011c. Google Groups. Welcome to Google Groups. Available at: <http://groups.google.com/googlegroups/overview.html> [Accessed June 30, 2011].
- Anon, Google Trends. Available at: <http://www.google.com/trends> [Accessed June 28, 2011a].
- Anon, 2011d. Ning. Ning – Sign In. Available at: <https://www.ning.com/main/signin> [Accessed July 1, 2011].
- Anon, search-engine-optimization-starter-guide.pdf. Available at: [http://static.googleusercontent.com/external\\_content/untrusted\\_dlcp/www.google.com/en/webmasters/docs/search-engine-optimization-starter-guide.pdf](http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/webmasters/docs/search-engine-optimization-starter-guide.pdf) [Accessed June 29, 2011b].
- Anon, 2010a. Teachers Net. Teacher Mailrings: Teachers.net. Available at: <http://teachers.net/mailrings/> [Accessed July 1, 2011].
- Anon, 2011e. Trailmeme. Available at: <http://trailmeme.com/> [Accessed June 30, 2011].
- Anon, 2011f. Twitter. Twitter log in page. Available at: <http://twitter.com/> [Accessed July 1, 2011].
- Anon, 2010b. Wikispaces. Wiki's for Everyone – Wikispaces. Available at: <http://www.wikispaces.com/> [Accessed June 30, 2011].
- Anon, 2011g. WordPress.com. Get a Free Blog Here. Available at: <http://wordpress.com/> [Accessed June 30, 2011].
- Anon, 2011h. Yahoo! Search – Web Search. Yahoo. Available at: <http://www.altavista.com/> [Accessed June 30, 2011].

- Anon, 2011i. YouTube. Home Page. Available at: <http://www.youtube.com/> [Accessed June 30, 2011].
- Chi, E.H., Pirolli, P. & Lam, S.K., 2007. Aspects of augmented social cognition: Social information foraging and social search. In Proceedings of the 2nd international conference on Online communities and social computing. the 2nd international conference on Online communities and social computing. Beijing, China: Springer-Verlag, pp. 60–69. Available at: <http://dl.acm.org/citation.cfm?id=1784297.1784305>.
- Computer Education Society of Ireland, 2011. CESI-list – Google Groups. Cesi-List. Available at: <https://groups.google.com/forum/#!forum/cesi-list> [Accessed July 1, 2011].
- Cuffe, L., 2010a. Eleven Tools For Education | Just another WordPress.com site. Available at: <http://eleventoolsforeducation.wordpress.com/> [Accessed June 30, 2011].
- Cuffe, L., 2011a. ElevenToolsFromMalta – home. Available at: <http://eleventoolsfrommalta.wikispaces.com/> [Accessed June 30, 2011].
- Cuffe, L., 2011b. SubjectResourcesForIrishEd – Stats. SubjectResourcesForIrishEd. Available at: <http://subjectresourcesforirished.wikispaces.com/space/stats/overview> [Accessed July 1, 2011].
- Cuffe, L., 2010b. Tools For Social and Group Education. Available at: <http://socedtools.blogspot.com/> [Accessed June 30, 2011].
- Cuffe, L., 2011c. Tools for social or group teaching on Trailmeme. tools for social or group teaching. Available at: [http://trailmeme.com/trails/Tools\\_for\\_social\\_or\\_group\\_teaching](http://trailmeme.com/trails/Tools_for_social_or_group_teaching) [Accessed June 29, 2011].
- Cuffe, L., 2010c. Tools for Social Teaching and Learning | Google Groups. Available at: <http://groups.google.com/group/tools-for-social-teaching-and-learning> [Accessed June 30, 2011].
- Cuffe, L., 2010d. YouTube – Social and Group learning tools intro, Available at: <http://www.youtube.com/watch?v=fnWhMI9Qm74> [Accessed June 30, 2011].
- Cuffe, L., 2010e. Eleven Tools for Social and Group Teaching, Available at: <http://prezi.com/jqqrik-y5n0q/eleven-tools-for-social-and-group-teaching/> [Accessed May 15 2012].
- Evans, M.P., 2007. Analysing Google rankings through search engine optimization data. *Internet Research*, 17(1), pp.21–37.
- Fetterly, D., Manasse, M. & Najork, M., 2004. Spam, damn spam, and statistics. In Proceedings of the 7th International Workshop on the Web and Databases colocated with ACM SIGMOD/PODS 2004 – WebDB '04. the 7th International Workshop. Paris, France, p. 1.
- Fisher, R.A., 1930. *The genetical theory of natural selection*. 1st ed., Oxford, England: Clarendon Press.
- George Price, 1972. Fisher's "fundamental theorem" made clear. *Ann. Hum. Genet*, 36, pp.129–140. Available at: [Wiley online library](#)
- Hargadon, S., 2011. educationalwikis – Examples of educational wikis. Educational Wiki's. Available at: <http://educationalwikis.wikispaces.com/Examples+of+educational+wikis> [Accessed June 29, 2011].
- King, A.B., 2008. *Website optimization*, Sebastopol, CA: O'Reilly Media, Inc.
- Kumar, R. & Tomkins, A., 2009. A Characterization of online search behavior. *Bulletin of the Technical Committee on Data Engineering*, IEEE computer Society, 32(2), pp.3–11.
- Langville, A.N. & Meyer, C.D., 2006. *Google page rank and beyond*, Princeton, N.J.: Princeton University Press.
- Meiss, M., Menczer, F. & Vespignani, A., 2005. On the lack of typical behavior in the global Web traffic network. In Proceedings of the 14th international conference on World Wide Web – WWW '05. the 14th international conference. Chiba, Japan, p. 510.
- Sharratt, M. & Usoro, A., 2003. Understanding knowledge-sharing in online communities of practice. *Electronic Journal on Knowledge Management*, 1(2), pp.187–196.
- Silverman, D., 2000. *Doing qualitative research: a practical handbook*, London: SAGE.
- Stanford School of Medicine, 2011. Publicizing & Site Searchability – Frequently Asked Questions (FAQs) – Web Support & Development – Information Resources & Technology (IRT) – Stanford University School of Medicine – Stanford Medicine. Available at: <http://med.stanford.edu/irt/webauthor/references/publicizing.html> [Accessed June 23, 2011].

- Usono, Abel et al., 2007. Trust as an antecedent to knowledge sharing in virtual communities of practice. *Knowledge Management Research & Practice*, 5(3).
- Vassallo, E., 2010. Smart Solutions. Available at: <http://www.smartsolutionsmalta.com/> [Accessed June 29, 2011].
- Wenger, E., 2006. Communities of practice. Available at: <http://www.ewenger.com/theory/index.htm>.
- Wenger, E., 1998. *Communities of practice: learning, meaning, and identity* 1st ed., Cambridge England: Cambridge University Press.
- Wilson, R., 2011. Web Marketing Today. The Web Marketing Checklist: 37 Ways to Promote Your Site. Available at: <http://www.wilsonweb.com/articles/checklist.htm> [Accessed July 1, 2011].
- Zipf, G.K., 1949. *Human behavior and the principle of least effort: an introduction to human ecology*, Cambridge Mass: Addison-Wesley Press.