

A comparison of sources of links for academic Web impact factor calculations

Item Type	Journal article
Authors	Thelwall, Mike
Citation	Thelwall, M. (2002), "A comparison of sources of links for academic Web impact factor calculations", Journal of Documentation, Vol. 58 No. 1, pp. 66-78.
DOI	10.1108/00220410210425412
Publisher	MCB UP Ltd
Journal	Journal of Documentation
Download date	2025-05-16 23:49:03
Link to Item	http://hdl.handle.net/2436/3140

A Comparison of Sources of Links for Academic Web Impact Factor Calculations

Mike Thelwall¹

School of Computing and Information Technology
University of Wolverhampton, Wulfruna Street, Wolverhampton, WV1 1SB, UK.
Email: cm1993@wlv.ac.uk

Abstract

There has been much recent interest in extracting information from collections of web links. One tool that has been used is Ingwersen's Web Impact Factor. It has been demonstrated that several versions of this metric can produce results that correlate with research ratings of British universities showing that, despite being a measure of a purely Internet phenomenon, the results are susceptible to a wider interpretation. This paper addresses the question of which is the best possible domain to count backlinks from, if research is the focus of interest. WIFs for British universities calculated from several different source domains are compared, primarily the .edu, .ac.uk and .uk domains, and the entire web. The results show that all four areas produce WIFs that correlate strongly with research ratings, but that none produce incontestably superior figures. It was also found that the WIF was less able to differentiate in more homogenous subsets of universities, although positive results are still possible.

Introduction

There has been much recent interest analysing web pages and links, especially with search engines, despite the relatively unregulated and problematical nature of this information source (Larson, 1996; Ingwersen, 1998; Smith, 1999; Snyder & Rosenbaum, 1999; Egghe, 2000; Thelwall, 2000; Bar-Ilan, 2001; Björneborn & Ingwersen, 2001; Cronin, 2001; Thelwall, 2001a). One theme, for example, has been the counting of pages in one area of the web that link to those in another. The two major issues for any such backlink analyses are reliance upon commercial search engines for raw data and the lack of any quality control process for web pages comparable to the refereeing of journal articles. A far-sighted proposal was made by Ingwersen (1998) for a calculation, the Web Impact Factor (WIF) that measured the overall impact of a web space by counting backlink pages over large areas of the web. This has a close analogy with impact factors for journals, as, indeed, backlink analysis has with citation analysis generally. The WIF can be used directly to measure impact on the web, or indirectly to provide a metric that may correlate with important offline phenomena. In the latter case, there have been studies to assess web journal impact (Snyder & Rosenbaum, 1999) as well as the research standing of sets of individual British departments (Thomas & Willett, 2000), and of sets of universities. Whilst early results did not demonstrate significant association with existing measures, they did develop the methodology and identify numerous problems with the reliability and validity of the data (Smith, 1999; Thelwall, 2000; Bar-Ilan, 2001;

¹ Thelwall, M. (2002). A comparison of sources of links for academic Web Impact Factor calculations, *Journal of Documentation*, 58(1), 60-72.

Björneborn & Ingwersen, 2001; Thelwall, 2001a). It has since been demonstrated by Thelwall (2002) that a series of variations of the WIF metric do correlate with research ratings for British universities. The problem of the large number of university web pages unrelated to research was dealt with by manually separating out the apparently research-related pages for one of the calculations. WIFs based upon the resulting figures produced the highest correlation with a measure of overall university research quality. The backlink pages in this study were all from the UK academic web domain, but the question that motivates the present study is whether changing this to other domains, for example the top level web educational domain, is capable of producing better results. In particular, it would be valuable to know whether WIFs based upon other domains would obviate the need for the elaborate classification process for research and non-research pages. It seems, for example, a plausible hypothesis that international web links between academic institutions would naturally discriminate in favour of higher quality target pages and produce a better result than purely national links.

The Web Impact Factor has suffered some modifications since its introduction by Ingwersen. The versions used here will be derived from his external WIF. For a given university web site U and a chosen web space S , the WIF used is as follows.

$$\frac{\text{The number of pages in } S \text{ but not } U \text{ containing at least one link to a page in } U}{\text{The number of full-time equivalent academic members of staff working for the university associated with } U}$$

The WIF calculated with S being the whole of the web, as indexed by AltaVista, will be known as the External WIF. Other WIFs will be named after the domain name ending of the area of the web covered, for example in the .edu WIF, S will be the set of all web pages with the domain name part of their URL ending in .edu. The denominator of this calculation is essentially a measure of the size of the institution. Evidence from previous studies suggests that it is more appropriate than a page count for the domain, as used in the original version (Thelwall, 2001a; Thelwall, 2002). Of course, this modification only makes sense in the context of WIF calculations for a university web site. Those tested here will be from the 98 British universities for which a research rating is available, and excluding as a result the private University of Buckingham, and the distance learning based Open University. The rating used is derived from the 1996 official government Research Assessment Exercise (RAE) (Quality Assurance Agency, 2001), the most recently completed. It represents what is believed to be, in an international context, an unusually authoritative source of national university-level research quality quantification because of its derivation from an exercise that decides the destination of billions of pounds of research funding.

The primary question addressed in this paper is whether the external WIF and the .edu WIF associate more strongly with research ratings than the .ac.uk WIF for British universities. In order to set the results in context, WIFs will also be calculated for all the original generic Top Level Domains (gTLDs) of the Internet, as well as for British commercial web sites and the whole of the UK domain. A secondary issue will be to confirm that staff numbers form a better denominator for the WIF, at least for universities, and to assess the extent to which the results are still useful in the alternative calculation. The validity of the results will also be discussed.

Methodology

The figures required for the various WIF calculations will be obtained from AltaVista advanced queries. Although previous studies have shown the results of this search engine to be unreliable (Smith, 1999; Snyder & Rosenbaum, 1999; Thelwall, 2000; Thelwall, 2001a), more recent results have identified an improvement (Thelwall, 2001b; Thelwall, 2002). Search engines do not index the entire web, only the publicly accessible pages that can be found by following links from previously known pages, subject to limitations of storage space, and crawler design choices. In particular, they do not index restricted access sites and unknown pages not identifiably linked to by other known pages. Lawrence & Giles (1999) estimated that the major search engines index not more than 16% of web pages that are publicly accessible and are directly or indirectly linked to by the site home page. The fraction of the web indexed by any search engine should, therefore, be treated with suspicion, as should any derived results. This does not, however, sound the death knell for cybermetrics, rather it provides an additional challenge. A useful analogy is with the Science Citation Index. This does not cover the entire universe of scholarly publications: it omits books, monographs and many (presumably less well cited) journals. Despite this, the journal Impact Factor derived from it is widely used. Its flaws probably also make it frequently misunderstood and misused, which is also a danger for search engine derived data.

It has been discovered that, despite the partial search engine coverage of the web, AltaVista did appear to cover UK academic web sites, on average, better than a purpose built web crawler (Thelwall, 2002). This was probably due to the retention by AltaVista of the URLs of older web pages no longer linked to. This gives particular confidence in its results for links originating at this set of sites, but not for those from other areas targeted at the UK academic web. AltaVista was chosen for its relatively large coverage of the web and its provision of advanced search facilities to enable link counting and web site page counting, as described below. In 2001, Google appeared to have a much larger coverage of the web, but did not have a mechanism for counting pages on a site.

The number of web pages on each site will need to be obtained for one set of calculations. This can be obtained by a host count advanced query, as shown below for the University of Coventry, which has two commonly used domain name endings: .cov.ac.uk and .coventry.ac.uk.

```
host:.cov.ac.uk OR host:.coventry.ac.uk
```

This will return a count of the number of pages that are indexed by AltaVista and have domain names ending in either .cov.ac.uk or .coventry.ac.uk. This count will exclude all pages that are unavailable to AltaVista, for example due to being password protected, or by being on a special list of pages not to be crawled by robots. It will also exclude pages that AltaVista has found out about but decided not to index, perhaps due to use of the HTML frames feature, and pages that are not known by AltaVista and not linked to by any other known pages.

It is also possible to use an advanced query to count the number of pages in one area of the web indexed by AltaVista that contain a link to a second. The format of the request to count all links to a university web site from an area that excludes that domain, is as follows. The example shown is for counting the pages indexed by AltaVista on web

sites with domain names ending in .edu that contain a link to a page at the University of Wolverhampton.

(link:.wlv.ac.uk OR link:.wolverhampton.ac.uk) AND domain:.edu

This form of query will be used for each of the original seven Internet generic Top Level Domain (gTLD) areas (ICANN, 2001). The .com, .net and .org domain descriptions are approximations since their availability is formally unrestricted.

- .edu: educational web sites, mainly from the USA and Canada
- .mil: reserved exclusively for the United States Military
- .net: mainly network organisations
- .gov: reserved exclusively for the United States Government
- .com: mainly commercial organisations, probably dominated by North America
- .int: for organizations established by international treaties between governments or Internet infrastructure databases
- .org: mainly organisations not fitting into the above categories and not subscribing to a national domain

In order to count all pages external to a university web site that contain a link to it, it is necessary to use the link command but to exclude all pages on the host site. For the University of Ulster this would be as follows.

(link:.ulst.ac.uk OR link:.ulster.ac.uk) AND NOT host:.ulst.ac.uk AND NOT
host:.ulster.ac.uk

The most complex query is to count backlink pages from an area containing the university web site. The following example is for counting the pages indexed by AltaVista on web sites with domain names ending in .ac.uk, but not those on the University of Essex web site that contain a link to a page at the University of Essex.

(link:.essex.ac.uk OR link:.sx.ac.uk) AND host:.ac.uk AND NOT host:.sx.ac.uk AND
NOT host:.essex.ac.uk

The above queries were modified for universities with only one domain name in use, such as the University of Aston (.aston.ac.uk) and for those with three or more such as the University of Lincoln and Humberside (.ulh.ac.uk, .lincoln.ac.uk, .humber.ac.uk). Following an initial assessment of the results, it was discovered that the results returned for St Andrews University were incorrect. The cause was identified as AltaVista's internal algorithm, which was terminating the domain name at the dash symbol instead of the end. As a result of this, the link:.st-and.ac.uk command, for example, was returning links to pages with domain names containing "st". To test this, the query

host:.st-and.ac.uk AND NOT host:.ac.uk

was sent, which returned many pages containing "st" outside the ac.uk domain. This problem appears to be specific to St Andrews university, connected to the fact that "and" is a keyword. For example the query

host:st-andrews.ac.uk AND NOT .ac.uk

returned no matches, indicating that in this case it was not just searching for domain names containing “st”. Further testing with domain names containing hyphens for universities outside the UK failed to reproduce this behaviour, tending to confirm the impression that it would be confined to domain names with a hyphen followed by an recognized AltaVista keyword.

Academic staff numbers for each institution, to be used in some WIF calculations, were obtained for 1998 from Noble’s Financial Yearbook (1999) and the Higher Education Statistics Agency (2001). The WIFs will be compared with research ratings derived from the RAE (Mayfield University Consultants, 2000).

Results

Table 1 shows Pearson’s correlation coefficients measured between ten WIFs and RAE averages for 96 British universities.

Table 1. *Pearson’s correlation between WIFs and research assessment average for 96 British universities - academic staff numbers as denominator (* = significant at 5% level, ** = significant at 1% level, *** = significant at 0.1% level)*

Type of WIF	Total Link Pages Counted	Correlation
.ac.uk	378,945	0.71***
External	2,883,536	0.70***
.uk	611,811	0.69***
.edu	367,158	0.69***
.org	250,796	0.64***
.com	729,470	0.60***
.mil	1,771	0.53***
.gov	27,528	0.51***
.int	1,530	0.47***
.co.uk	162,054	0.44***
.net	159,602	0.23*

Table 2 shows Pearson’s correlation coefficients measured between ten WIFs and RAE averages for 95 British universities, using site page counts for each WIF denominator. The calculations exclude Glamorgan University, which had a domain count of only three pages in AltaVista, resulting in an enormous individual WIF calculation and much lower correlations overall. The low count for Glamorgan does not reflect its real web site size, which may be due to the AltaVista being unable to meaningfully index it due to its design around the HTML frames feature. It can be seen that the correlations are much lower than those for table 1, confirming that staff numbers are a better choice for WIF denominators.

Table 2. *Pearson's correlation between WIFs and research assessment average for 95 British universities - page numbers as denominator (* = significant at 5% level, ** = significant at 1% level, *** = significant at 0.1% level)*

Type of WIF	Correlation
.edu	0.46***
External	0.31**
.int	0.31**
.mil	0.29**
.org	0.21*
.com	0.19
.gov	0.16
.ac.uk	0.12
.net	0.11
.uk	-0.01
.co.uk	-0.04

Figures 1 to 3 show WIFs against RAE average for the main three types considered. One general trend is that the new universities, all of which have an RAE average of 1.4 or below, are grouped together with a low WIF compared to the rest. The traditional universities are much more widely scattered in absolute value. Linear trends are also clear in the graphs, as is the presence of wide scattering in WIF values, even for institutions with similar research averages. Figure 4 shows that, to scale, the new universities also exhibit wide variations in WIF values.

Figure 1. *A graph to illustrate the relationship between external WIFs, with academic staff numbers denominator, and RAE averages for 96 British universities*

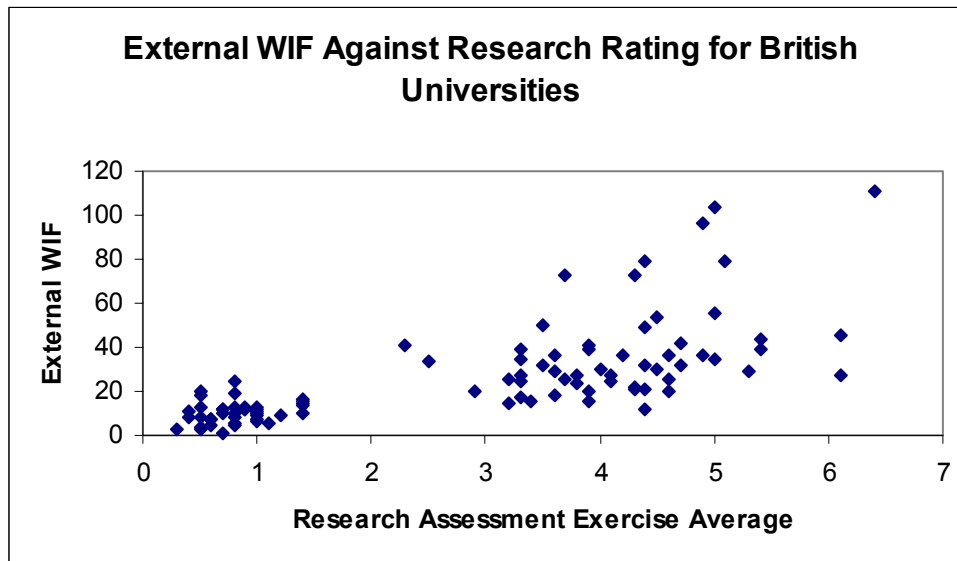


Figure 2. *A graph to illustrate the relationship between .edu WIFs, with academic staff numbers denominator, and RAE Averages for 96 British universities*

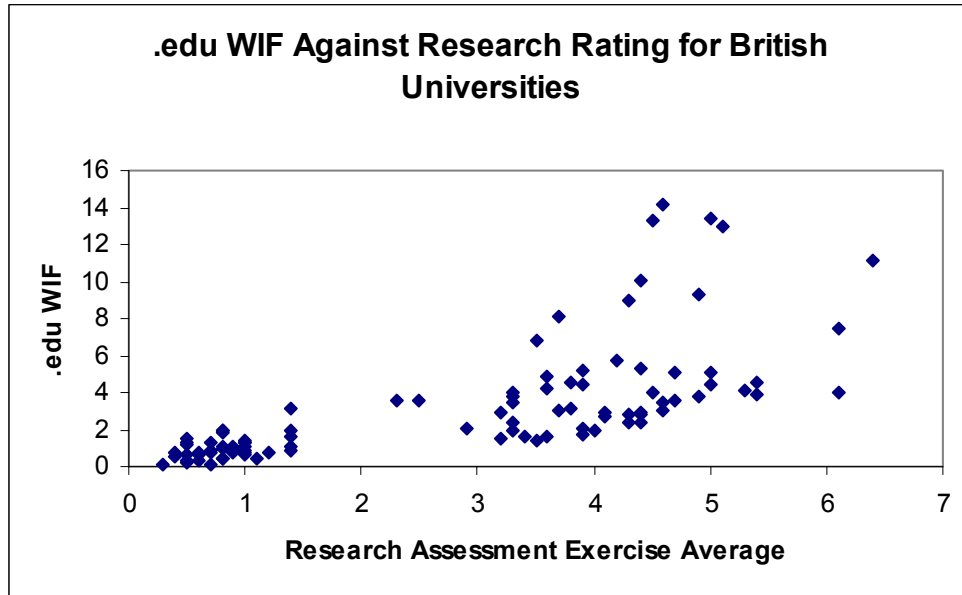


Figure 3. *A graph to illustrate the relationship between .ac.uk WIFs excluding links from the host institution, with academic staff numbers denominator, and RAE averages for 96 British universities*

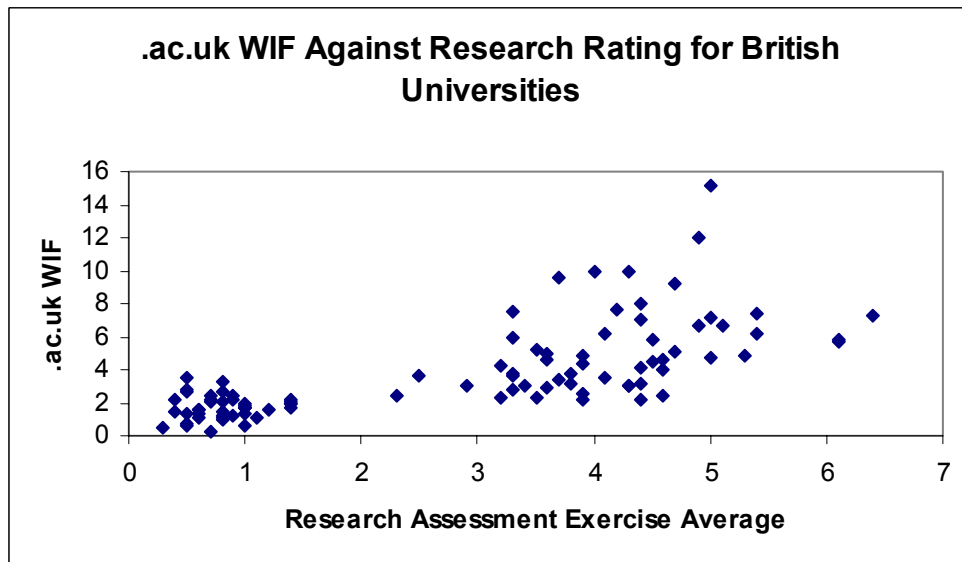
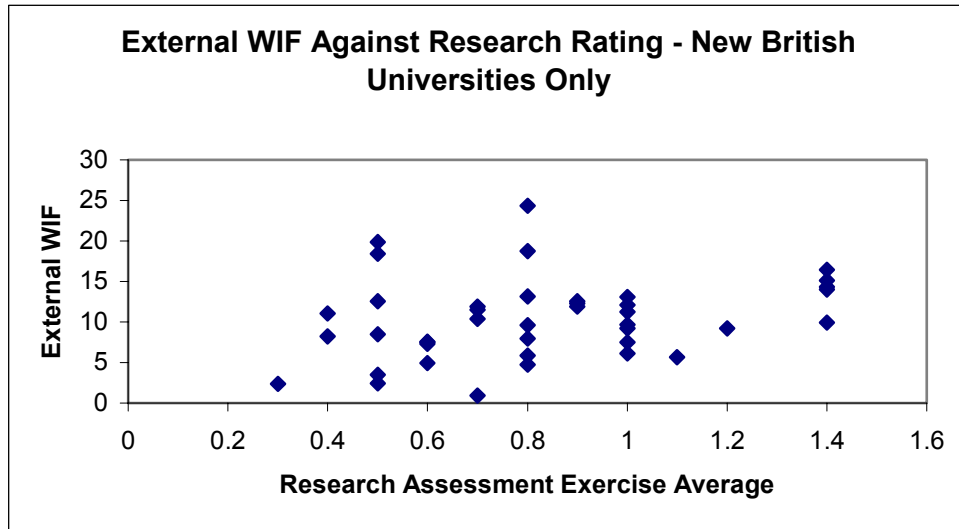


Figure 4. A graph to illustrate the relationship between external WIFs, with academic staff numbers denominator, and RAE averages for the 41 British new universities



In the light of the figures 1 to 4, a possible interpretation of the high correlation coefficients reported for most WIFs is their ability to discriminate between universities with a relatively low level of research and those with one much higher. The results shown thus far do not conclusively demonstrate any ability to associate with research within groups of more similar research profiles. This would be an issue if the approach was extended to other countries with a smaller range of research quality in universities, as may well be currently the case in the majority of Europe. To test the ability of the WIF to discriminate in a smaller spectrum of universities, correlations were calculated for the new universities separately from the traditional, with the results shown in table 3. It can be seen that the correlations are much lower, but that some significance is still evident. For the new universities, only the .edu WIF shows a strong association, although for the traditional group several correlations are nevertheless significant.

Table 3. *Pearson's correlation between WIFs and Research Assessment Average for 41 British new universities and for 55 older universities - academic staff numbers as denominator (* = significant at 5% level, ** = significant at 1% level, *** = significant at 0.1% level)*

Type of WIF	Correlation for New Universities	Correlation for Older Universities
.org	0.30	0.48***
.edu	0.46**	0.43***
External	0.27	0.42**
.ac.uk	0.09	0.40**
.uk	0.07	0.31*
.mil	0.22	0.23
.com	0.26	0.19
.gov	0.39*	0.15
.int	0.23	0.11
.net	0.09	0.06
.co.uk	0.20	0.01

Discussion

The findings for the British university WIF calculations are summarised below.

- The .edu, External and .uk WIFs all correlate at a similar level with research to .ac.uk WIF for British universities.
- The .org WIF also produces a high correlation with research for British universities.
- The associations are much stronger when taken over the full spectrum of British universities than when restricted to those with similar research ratings.
- The .edu WIF associates most strongly with research ratings for the group of new universities.
- At all levels of research rating there was a great variation in WIF for similar universities.
- Academic full time equivalent staff numbers are confirmed as a better choice for WIF denominator than web site page counts. If page counts are used for domains, then a minimum size requirement may need to be imposed in order to weed out sites sparsely indexed by AltaVista.

It would be valuable to be able to construct plausible explanations for unusual individual WIF values in order to develop an understanding of underlying sources of variation. The lowest WIF scores came from the University of Central England, which hosts almost no research on its web site. It appears to have a small site publishing authorised pages only and consequently has attracted few recorded links. The three unusually high scores for new university external WIFs came from Napier (RAE 0.5, external WIF 19.9), Wolverhampton (RAE 0.5, external WIF 18.4) and Staffordshire (RAE 0.8, external WIF 24.3). These are all universities with a higher than average proportion of staff working in computing-related departments, as recorded in the 1998 HESA statistics (Higher Education Statistics Agency, 2001). This could be an explanation, under the assumption that computing departments will make more use of the web and, therefore, that the web publications of computing academics are more likely to attract links. The four traditional universities with relatively high .edu WIFs were: Leeds (RAE 4.5, .edu WIF 13.3);

Durham (RAE 4.6, .edu WIF 14.2); York (RAE 5, .edu WIF 13.4); and Edinburgh (RAE 5.1, .edu WIF 13.0) do not fit this pattern, however, with Leeds and Durham having below average proportions of academic computing staff. An investigation of the source of Durham's links revealed that 12,784 of the 39,535 came from a single source - the online archive of California (<http://sunsite2.berkeley.edu/oac/>). AltaVista had apparently indexed over ten thousand different URLs in the site that Berkeley's server was redirecting to a link list page that included a link to Durham's library. A similar investigation from Leeds suggested that many of the backlink pages were the result of the use of Nikos Drakos' LaTeX2HTML program. This program converts documents typeset in LaTeX into HTML files, putting a credit link to the author in the web pages produced. It was not possible to check this directly since AltaVista did not index the word LaTeX2HTML fully, but a search for pages outside Leeds linking to the university web site that contained both of the telltale phrases "Nikos Drakos" and "this document was generated using the" indicated that 58,279 of the total 109,433 backlink pages came from this source. This single project has, therefore, had a large impact on the total link count primarily because of its nature rather than its (undisputed) research value. It is clear, then, that technical problems and individual Internet projects can have a large impact on WIF calculations.

It is not known for certain whether the results would have been different if another search engine had been used, and if the results are artefacts of AltaVista's coverage or search engine coverage in general rather than being 'genuine' web phenomena. The latter question is difficult to tackle because of the dynamic nature of the web and the need to develop a systematic approach to issues such as site mirroring and publicly accessible pages that are not linked to and are not found by crawlers. The significant correlations discovered are evidence that there is sense in continuing to use search engines for data collection of this type, and that it is meaningful to restrict interest to the collection of web pages that they are capable of indexing. If Google, for example, had allowed the queries necessary for the calculations performed then it would have been interesting to compare the results. The results may well have been significantly different for domains producing few link pages, such as .int and .mil, which did not show any links at all to some universities, but the results for higher scoring domains should be inherently more stable, at least in relative size. The lowest count for the .edu domain was 97 (mean 3,826), for ac.uk: 291 (mean 4,072), for .org: 96 (mean 2,618), and the all external links: 1,241 (mean 30,079). Significant differences in the magnitude of individual counts could arise from policy decisions concerning the scope and coverage of any particular engine, for example if one area of the web were given preference over another. There is, in fact, some weak evidence of this happening with AltaVista, since its coverage of the web (as described above) appeared to be below 16% in 1999 (Lawrence & Giles, 1999), whereas in 2000 its coverage of UK academic web sites seemed to be greater than that which would have been covered by Lawrence and Giles's crawler (Thelwall, 2002). If this greater coverage is a reality, then it could be explained by the academic web being, in general, better interlinked and older than the commercial web. If age is an important factor, then the newer Google could indeed produce results of a significantly different relative magnitude between, say, the ac.uk domain and the .org domain. This is, nevertheless, still not an argument for a likelihood of significantly different correlation coefficients within a single top level or academic domain because significant changes in

relative magnitude of counts to sites would have to occur. In effectively composite calculations, such as for the External WIF, a relatively greater number of .com pages, say, than ac.uk pages could bias the calculation towards more commercially oriented universities. For the ac.uk WIF or the .com WIF individually, an increase in the results would not necessarily mean differing relative magnitudes of individual WIFs that would be enough to significantly change correlation coefficients. The key question for an individual source domain is whether the selection of the web by one search engine over another would produce a bias in the results for one or a group of universities. This is possible for individual universities that create web sites that are mirrored to external sites and contains many links to the host institution. In this case, differing policies with regard to identifying and indexing such duplicate collections could make a real difference to individual WIFs. It seems unlikely, however, that coverage phenomena should produce a significant bias in the results for more than three or four UK universities.

The validity of the non-web data used is not beyond question, and merits some discussion. The RAE data is now over five years old, and it will be interesting to see whether figures based upon the 2001 exercise differ significantly from the earlier ones. The outcome of the assessment in any particular department is not just the result of the research, but also of a tactical decision about the likely outcome in relation to the number of staff submitted and funding levels. For example, two identical departments may achieve different ratings because one has submitted more research active staff, with a lower overall average score. This kind of problem with the data is, unfortunately, unavoidable, but it is believed that the overall result across an institution is still representative of its research profile. It has also been raised as an issue with the author that the methods of counting academic staff numbers may well vary between institutions, particularly concerning the decision about which members of staff may be classed as academic. It seems that, for practical reasons, this must be tolerated, although it may be the case that variations from this source will be invisible in the context of much larger variations from web links.

Conclusions

The results show that several WIFs, including the .edu, external, .ac.uk, .uk and .org versions, produce results that correlate significantly with research level across the spectrum of British universities, although the association is reduced or disappears when subsets of similar institutions are compared. The variable results for the different calculations have not thrown into relief any one of the sources as demonstrably superior to the .ac.uk domain, and so the primary finding of this paper is negative the sense of finding a replacement for the .ac.uk domain for UK university WIFs. The discovery that the .edu domain is not more discriminatory than the ac.uk domain is perhaps surprising given that a key criteria for higher quality in research is that it is of international significance.

It is encouraging that a variety of WIFs from different sources produce significant correlations with research ratings. This should make it easier to transfer the results to other domains, particularly in the case of countries that do not have designated academic national sub-domains, as is the case in the majority of Europe. Since the results are for Britain alone, however, more countries need to be studied in order to develop a level of confidence in the international applicability of the measure, and to decide which WIF

versions are generally the most reliable. In this context, the .org WIF was an unexpected source of association with research, coming from a domain that does not contain university web sites, and the reason for this would be an interesting source of further research.

The WIF calculations produced here are clearly not accurate enough to provide any kind of reliable substitute measure of research at the individual university level. In particular the results are less reliable than citation counting based exercises, which have been proposed as the basis for an alternative to the RAE in Britain (Oppenheim, 1997; Warner, 2000; Holmes & Oppenheim, 2001). The analysis of individual results was, however, able to identify sources of variation, but it is possible that had it been pursued systematically, improved results could have been obtained. A future direction for research is, therefore, to attempt to define rigorous data cleansing procedures with the objective of increasing the reliability of the WIF for individual institutions without having to resort to the classification of the research content individual pages.

Acknowledgements

The author would like to thank the referees for their helpful suggestions.

References

- Bar-Ilan, J. (2001). Data collection methods on the web for informetric purposes - a review and analysis. *Scientometrics*, 50(1), 7-32.
- Bjørneborn, L. & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50(1), 65-82.
- Cronin, B. (2001). Bibliometrics and Beyond: Some thoughts on web-based citation analysis. *Journal of Information Science*, 27(1), 1-7.
- Egghe, L. (2000). New informetric aspects of the Internet: some reflections - many problems. *Journal of Information Science*, 26(5), 329-335.
- Higher Education Statistics Agency (2001). <http://www.hesa.ac.uk/> (visited 27 March 2001).
- Holmes, A. & Oppenheim, C. (2001). Use of citation analysis to predict the outcome of the 2001 RAE for Unit of Assessment 61: Library and Information Management. *Information Research*, 6 (2).
- ICANN (2001). Domain Name Supporting Organization of ICANN ICANN/DNSO Constituency Groups. <http://www.dns0.org/constituency/gtld/gtld.html> (visited 29 March 2001).
- Ingwersen, P. (1998). The calculation of Web Impact Factors. *Journal of Documentation*, 54(2), 236-243.
- Larson, R. R. (1996). Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace. ASIS 96. <http://sherlock.berkeley.edu/asis96/asis96.html> (visited 4 August 2001).
- Lawrence, S. & Giles, C. L. (1999). Accessibility of information on the web. *Nature*, 400, 107-109.
- Mayfield University Consultants (2000). League Tables 2000. *The Times Higher Education Supplement*, April 14, II-III.
- The Noble Publishing Company (1999). *Noble's Higher Education Financial Yearbook 1999*, Noble.

- Oppenheim, C. (1997). The correlation between citation counts and the 1992 research assessment exercise ratings for British research in genetics, anatomy and archaeology. *Journal of Documentation*, 53, 477-487.
- Quality Assurance Agency, (2001). <http://www.qaa.ac.uk> (visited 28 March 2001).
- Smith, A. G. (1999). A tale of two web spaces: comparing sites using Web Impact Factors. *Journal of Documentation*, 55(5), 577-592.
- Snyder, H. & Rosenbaum, H. (1999). Can search engines be used for web-link analysis? A critical review. *Journal of Documentation*, 55(4), 375-384.
- Thelwall, M. (2000). Web Impact Factors and search engine coverage. *Journal of Documentation*, 56(2), 185-189.
- Thelwall, M. (2001a). Results from a Web Impact Factor crawler. *Journal of Documentation*, 57(2), 177-191.
- Thelwall, M. (2001b). The responsiveness of search engine indexes. *Cybermetrics*, 5(1),. <http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html> (visited 28 March 2001)
- Thelwall, M. (2002, to appear). Extracting macroscopic information from web links. *Journal of the American Society of Information Science and Technology*.
- Thomas, O. & Willet, P. (2000). Webometric analysis of departments of Librarianship and information science. *Journal of Information Science*, 26(6), 421-428.
- Warner, J. (2000). A critical review of the application of citation studies to the Research Assessment Exercises. *Journal of Information Science*, 26(6), 453-460.