

Invited plenary talk
“WISER Webometrics”

Mike Thelwall

School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK. E-mail: m.thelwall@wlv.ac.uk
Tel: +44 1902 321470 Fax: +44 1902 321478

Webometrics in information science is currently dominated by link analysis and strongly influenced by citation analysis because it is typically applied to scientific documents. The WISER project was an EU-funded consortium tasked with exploring the potential of the internet to yield a new generation of scientometric indicators. This talk describes some of the findings of this project and discusses the wider potential for link analysis research in scientometrics.

The WISER consortium was lead by Andrea Scharnhorst and Paul Wouters, now of the Virtual Knowledge Studio (VKS - KNAW) in Amsterdam, and the other members were Isidro Aguillo's team in Madrid (CINDOC), Hiltrun Kretchmer, working from Amsterdam, and Sylvan Katz (SPRU) and Mike Thelwall and Viv Cothey (Wolverhampton) in the UK. The experience of the team included information science, sociology, science policy, indicator development and computer science. The list of publications below gives an idea of the wide range of findings produced by the consortium, and the following generic framework for link

analysis in social sciences research, developed in the WISER project, illustrates one of the outcomes that others may find useful (Thelwall, 2004).

- 1) Formulate an appropriate research question, taking into account existing knowledge of web structure.
- 2) Conduct a pilot study.
- 3) Identify web pages or sites that are appropriate to address a research question.
- 4) Collect link data from a commercial search engine or a personal crawler, taking appropriate safeguards to ensure that the results obtained are accurate.
- 5) Apply data cleansing techniques to the links, if possible, and select an appropriate counting method.
- 6) Partially validate the link count results through correlation tests.
- 7) Partially validate the interpretation of the results through a link classification exercise.
- 8) Report results with an interpretation consistent with link classification exercise, including either a detailed description of the classification or exemplars to illustrate the categories.
- 9) Report the limitations of the study and parameters used in data collection and processing (stages 3 to 5).

Although the WISER project has demonstrated the potential of link analysis, there are many potential future research directions, including the following.

- *Investigations into why links are created, particularly in academic contexts.*
- *Time series analyses.*
- *Applying social network analysis measures to information collections.*
- *Supporting wider social sciences research.*
- *Developing new visualisation methods.*

Wiser Publications

Books

Thelwall, M. (2004). *Link Analysis: An Information Science Approach*. San Diego, CA: Academic Press.

Edited special issues

1. Scharnhorst, A., van den Besselaar, P., Wouters, P. (Eds.) (2006). What does the Web represent? From virtual ethnography to web indicators. Special issue *Cybermetrics* (2006).
2. Katz, S. (Ed) (2006) Web indicators for innovation systems. Special issue *Research Evaluation*.

International and national peer reviewed journals

1. Cothey, V. (2004). Web-crawling reliability. *Journal of the American Society for Information Science and Technology*, 55(14), 1228-1238.
2. Ortega Priego, J. L. (2003) A Vector Space Model as a methodological approach to the Triple Helix dimensionality: A comparative study of Biology and Biomedicine Centres of two European National Research Councils from a Webometric view. *Scientometrics*, 58(2), 429-443.
3. Thelwall, M. & Price, E. (2003). Disciplinary differences in academic web presence - A statistical study of the UK. *Libri*, 53(4), 242-253
4. Thelwall, M., & Aguillo, I. (2003). La salud de las Web universitarias españolas. *Revista Española de Documentación Científica*, 26(3), 291-305.

5. Thelwall, M. (2004, to appear). Weak benchmarking indicators for formative and semi-evaluative assessment of research. *Research Evaluation*.
6. Thelwall, M., Vaughan, L., Cothey, V., Li, X. & Smith, A. (2003). Which academic subjects have most online impact? A pilot study and a new classification process. *Online Information Review* 27(5), 333-343.
7. Thelwall, M., Harries, G., & Wilkinson, D. (2003). Why do Web sites from different academic subjects interlink? *Journal of Information Science*, 29(6), 445-463.
8. Thelwall, M., Harries, G. (2004). Do better scholars' web publications have significantly higher online impact?, *JASIST* 55(2), 149-15
9. Wilkinson, D., Thelwall, M. & Li, X. (2003). Exploiting hyperlinks to study academic Web use. *Social Science Computer Review*, 21(3), 340-351.
10. Wilkinson, D., Harries, G., Thelwall, M. & Price, E. (2003). Motivations for academic Web site interlinking: Evidence for the Web as a novel source of information on informal scholarly communication. *Journal of Information Science*, 29(1), 59-66.
11. Cothey, V. & H. Kretschmer (2004): Does the link structure of the web provide evidence of a collaborative hypertext? *Journal of Information Management and Scientometrics*. Vol. 1, No. 2
12. Harries, G., Wilkinson, D., Price, E., Fairclough, R. & Thelwall, M. (2004). Hyperlinks as a data source for science mapping, *Journal of Information Science*, 30(5), 436-447.
13. Kretschmer, H (2004): Author productivity and geodesic distance in bibliographic co-authorship networks, and visibility on the Web. *Scientometrics*. Vol.60, No.3, 409-420
14. Kretschmer, H. & I.F. Aguillo (2004): Visibility of collaboration on the Web. *Scientometrics*. Vol. 61, No. 3, (2004), 405-426.
15. Kretschmer, H.& M Thelwall (2004): From Librametry to Webometrics. *Journal of Information Management and Scientometrics*, Vol. 1, No. 1, 1-7
16. Tang, R. & Thelwall, M. (2004). Patterns of national and international web inlinks to US academic departments: An analysis of disciplinary variations. *Scientometrics* (ISSI 2003 special issue), 60(3), 475-485.
17. Thelwall, M., & Vaughan, L. (2004). A fair history of the Web? Examining country balance in the Internet Archive. *Library & Information Science Research*, 26(2), 162-176.
18. Thelwall, M. & Harries, G. (2004). Can personal web pages that link to universities yield information about the wider dissemination of research? *Journal of Information Science*, 30(3), 243-256.
19. Thelwall, M. (2004). Can the Web give useful information about commercial uses of scientific research? *Online Information Review*, 28(2), 120-130.
20. Thelwall, M. (2004). Weak benchmarking indicators for formative and semi-evaluative assessment of research. *Research Evaluation*, 13(1), 63-68.
21. Thelwall, M. & Wilkinson, D. (2004). Finding similar academic Web sites with links, bibliometric couplings and colinks. *Information Processing & Management*, 40(3), 515-526.
22. Thelwall, M. (2004). Methods for reporting on the targets of links from national systems of university Web sites. *Information Processing & Management*, 40(1), 125-144.
23. Thelwall, M. & Vaughan, L (2004). New versions of PageRank employing alternative Web document models. *ASLIB Proceedings*, 56(1), 24-33.
24. Thelwall, M., & Harries, G. (2004). Do better scholars' web publications have significantly higher online impact? *Journal of the American Society for Information Science and Technology*, 55(2), 149-159.
25. Thelwall, M. & Vaughan, L. (2004). Webometrics: An introduction to the special issue *Journal of the American Society for Information Science and Technology*. 55(14), 1213-1215.
26. Vaughan, L. & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes. *Information Processing & Management*, 40(4), 693-707.
27. Scharnhorst, A., Wouters, P. (2006) Webindicators. A new generation of science and technology indicators. *Cybermetrics* Vol. 10, Iss. 1.
28. Cothey, V., Aguillo, I. and Arroyo, N. (2006) Operationalising "Websites": lexically, semantically or topologically? *Cybermetrics* Vol. 10, Iss. 1.
29. Thelwall, M. & Stuart, D. (2006, to appear). Web crawling ethics revisited: Cost, privacy and denial of service. *Journal of the American Society for Information Science and Technology*.
30. Thelwall, M., Barjak, F. & Kretschmer, H. (2006, to appear). Web links and gender in science: An exploratory analysis. *Scientometrics*, 67(3)
31. Scharnhorst, A., Thelwall, M. (2005) Citation and Hyperlink Networks. *Current Science* Vol. 89, No. 9, pp. 1518-1523.
32. Thoms, L. & Thelwall, M. (2005). Academic home pages: Reconstruction of the self, *First Monday*, 10(12), http://www.firstmonday.org/issues/issue10_12/thoms/

33. Thelwall, M. (2006). Interpreting social science link analysis research: A theoretical framework. *Journal of the American Society for Information Science and Technology*, 57(1), 60-68.
34. Thelwall, M. (2005). Creating and using web corpora, *International Journal of Corpus Linguistics*, 10(4), 517-541.
35. Kretschmer, H. & Aguillo, I. F. (2005). New indicators for gender studies in Web networks. *Information Processing & Management*, 41 (6): 1481-1494
36. Aguillo, I. F.; Granadino, B.; Ortega, J.L. & Prieto, J.A. (2005). What the Internet says about Science. *The Scientist*, 19(14):10, Jul. 18, 2005
37. Aguillo, I. F.; Granadino, B. & Llamas, G. (2005). Posicionamiento en el Web del sector académico iberoamericano[Web positioning of the Iberoamerican academia sector]. *Interiencia*, 30(12):1-5
38. Aguillo, I. F.; Granadino, B. (2005). Análisis de la presencia institucional del CSIC en el Web [Análisis de the presence of the CSIC in the Web]. *Arbor*, 705 (Septiembre 2004):111-139.
39. Aguillo, I. F. (2005). Evaluación de la presencia en el Web de la UNAM y otras universidades mexicanas [Evaluation of the Web presence of the UNAM and other Mexican universities]. *Revista Digital Universitaria*. 10 de agosto 2005, Vol. 6, No. 8. <http://www.revista.unam.mx/vol.6/num8/art83/int83.htm>
40. Cueto, L.; Soler, Ch.; Entremozaga, R.; Martínez, E.; Carrasco, A.; Aguillo, I.; Fernández, E.; Rodríguez, M.; Barrantes, D.; Espadas J.; Ferreras R.; Albert, J.; Sanz de las Heras, J. (2005). Técnicas de análisis de posicionamiento.[Methods of web positioning analysis]. *El Profesional de la Información*, 2005, 14(1): 14-20.
41. Ortega, J. L. (2005) Análisis de sesiones de la web del Cindoc: una aproximación a la minería de uso web [Análisis of the log files of the Cindoc website: A web data mining study]. *El profesional de la información*, 14(3): 190-198
42. Aguillo, I. F. (2005). Observatorio: Evaluación de revistas electrónicas [Observatory: Evaluation of electronic journals]. *El Profesional de la Información*, Septiembre-octubre 2005, 14(5):324-325.
43. Silió, T. (2005). Los fundamentos tecnológicos del acceso abierto: Open Archives Initiative y Open Archival Information System [Technological aspects of the Open Access: Open Archives Initiative and Open Archival Information System]. *El Profesional de la Información*, Septiembre 2005, 14(5):365-380.
44. Aguillo, I. F. (2005). Indicadores de contenidos para la web académica iberoamericana [Content indicators for Iberoamerican academic webs]. *BiD: textos universitaris de biblioteconomia i documentació*, diciembre, núm. 15. <http://www2.ub.edu/bid/consulta_articulos.php?fichero=15aguil2.htm>
45. Kretschmer, H. & I.F. Aguillo (2005) New Indicators for Gender Studies in Web Networks. *Information Processing & Management*. Volume 41, Issue 6, December 2005, 1481-1494
46. Kretschmer, H., U. Kretschmer, T. Kretschmer (2005) Reflection of Co-authorship Networks in the Web: Web Hyperlinks versus Web Visibility Rates. (accepted for publication in *Scientometrics*)
47. Kretschmer, H., U. Hoffman & T. Kretschmer (2005) Reflection of Collaboration Structures between German Immunology Institutions in the Web and Gender Visibility (submitted for publication in *Research Evaluation*)
48. Aguillo, I. & H. Kretschmer (2005) Creation and Identification of Personal and Institutional Web-pages (accepted for publication in *Journal of Information Management and Scientometrics*)
49. Kretschmer, H. (2005) Web indicators of evaluation and collaboration tested in Europe and in India (accepted for publication in *Journal of Information Management and Scientometrics*).

Chapters, contributions to monographs, proceedings

1. Kretschmer, H.(2003): Author productivity and Erdős distances in co-authorship and in web link networks. In: Jiang Guohua, Ronald Rousseau & Wu Yishan (Eds.), Proceedings of the 9th International Conference on Scientometrics and Informetrics, ISSI 2003, Dalian University of Technology Press: Dalian, 2003, pp. 393-400
2. Kretschmer, H. and M. Thelwall (2003): The development of information professionals: The European perspective: The way from librametry to webometrics. In: A. Amudhavalli (Ed), Proceedings of the MALA Platinum Jubilee Celebrations. National Seminar on Information Professionals for the Digital Era, Madras, India, January 29-30, 2003. EFEX: Chennai, 13-25

3. Tang, R. & Thelwall, M. (2003). Patterns of international and national Web inlinks to US university departments: A Webometric analysis of disciplinary specificity. In: Jiang Guohua, Ronald Rousseau & Wu Yishan (Eds.), *Proceedings of the 9th International Conference on Scientometrics and Informetrics, ISSI 2003*, Dailan University of Technology Press, Dailan, China, pp. 312-320.
4. Cothey, V. & H. Kretschmer (2004): Does the link structure of the web provide evidence of a collaborative hypertext? In: Hildrun Kretschmer, Yogendra Singh. & Ramesh Kundra (Eds.). *Proceedings of the International Workshop on Webometrics, Informetrics and Scientometrics & 5th COLLNET Meeting, 2-5 March 2004*. Central Library, Indian Institute of Technology Roorkee: Roorkee-247667, India, 82-90
5. Kretschmer, H. & I. Aguillo (2004): COLLNET Part I.: Visibility of collaboration on the web. In: Hildrun Kretschmer, Yogendra Singh. & Ramesh Kundra (Eds.). *Proceedings of the International Workshop on Webometrics, Informetrics and Scientometrics & 5th COLLNET Meeting, 2-5 March 2004*. Central Library, Indian Institute of Technology Roorkee: Roorkee-247667, India, 150-172
6. Kretschmer, H. & I. Aguillo (2004): COLLNET Part II: New indicators of gender visibility in web networks. In: Hildrun Kretschmer, Yogendra Singh. & Ramesh Kundra (Eds.). *Proceedings of the International Workshop on Webometrics, Informetrics and Scientometrics & 5th COLLNET Meeting, 2-5 March 2004*. Central Library, Indian Institute of Technology Roorkee: Roorkee-247667, India, 173-184
7. Kretschmer, H & T. Kretschmer (2004): Comparison of rules in bibliographic and in web networks. In: T.A.V. Murthy, S.M. Salgar, G. Makhdumi, P. Pichappan, Y. Patel and J. K. Vijayakumar (Eds.). *Proceedings of the Second International Caliber 2004: Road Map to New Generation of Libraries Using Emerging Technologies*. February 11-13, 2004, New Delhi, India. INFLIBNET Centre: Ahmedabad, 470-486.
8. Cothey V. (2005). Some preliminary results from a link-crawl of the European Union research area Web. In Ingwersen P. and Larsen, B (eds.) *Proceedings of the tenth international conference of the International Society for Scientometrics and Informetrics*, pp 212-220, Stockholm, July 24-28, 2005. Karolinska U.P.
9. Thelwall, M. & Wouters, P. (2005). What's the deal with the web/Blogs/the next big technology: A key role for information science in e-social science research? CoLIS 2005, *Lecture Notes in Computer Science 3507*, 187-199.
10. Kretschmer, H., U. Kretschmer, T. Kretschmer (2005) Reflection of Co-authorship Networks in the Web: Web Hyperlinks versus Web Visibility Rates. In: *Proceedings of the 5th Triple Helix Conference*, May 18-21, 2005, Turin, Italy (CD-ROM)
11. Kretschmer, H., U. Kretschmer, T. Kretschmer: Visibility of Collaboration between Immunology Institutions on the Web Including Aspects of Gender Studies. In: Peter Ingwersen & Birger Larsen (Eds.). *Proceedings of the 10th ISSI International Conference on Scientometrics and Informetrics*, July 24-28, 2005, Stockholm, Sweden, Volume 2. Published by Karolinska University Press: Stockholm, 2005, 750-760.
12. Katz, J.S. (2005). Scale Independent Bibliometric Indicators. *Measurement: Interdisciplinary Research and Perspectives* (Forthcoming).