Linking to a bilingual Web space

K<small>IM</small> H<small>OLMBERG</small>[1]

1 kim.holmberg@abo.fi, Åbo Akademi University, Department of Information Studies, Fänriksgatan 3, 20500 Åbo, FINLAND

Abstract

The goal of this research was to study whether there is a significant correlation between how bilingual municipal Web sites receive links to their two domains in two languages and the extent of these languages among the population in the municipalities. It was shown that the domain in the majority language receive the majority of the links. Statistically significant correlations were found, suggesting that those creating the links are well aware of the language situation in the municipalities and use that knowledge when creating the links. Search engine Yahoo was found to give the most accurate results of link counts to the researched bilingual web space. Yahoo can therefore be recommended for webometric studies where language is a factor.

1. Introduction

Keeping in mind the nature of the Web, where anyone can claim their own little corner and publish anything on it without any censorship or quality control, it is remarkable that results from webometric studies can correlate so strongly with other information sources of known value. Webometric studies have successfully used the amount of links as a measure of visibility [1, 2] and impact of a collection of Web sites [3, 4] or different academic subjects [5]: the more inlinks a site receives, the more it has been recognized and linked to by others. More inlinks mean that a site or a page is easier to find, partly

because it is easier to find a link when there are more links to be found but also because we can assume that at least some search engines use the number of inlinks as part of their ranking algorithms [6]. Links have been used to trace patterns of informal scientific communication [7] and they have been found to correlate with research ratings in universities [4, 8, 9, 10, 11]. Geographic patterns have also been shown to have an impact on linking [12, 13]. Vaughan and Thelwall [14] found that inlink counts to a Web site correlate with the age of the site. Older sites with richer content are more visible on the Web and therefore receive more inlinks. This seems to indicate a cumulative effect where rich get richer. Barjak and Thelwall [15] studied different characteristics of research teams that might explain the inlink counts to their Web sites. They discovered that the size of both the research team and their Web presence correlated with inlink counts, but that research productivity did not. Some earlier webometric studies have found strong correlations between inlink counts and other offline measures, showing how different offline aspects influence linking, but little is known about how language and especially domains in different languages attract links.

This research will increase knowledge about the role of language in linking. It will be studied how bilingual municipalities with two domains in two languages receive inlinks and whether there is a correlation between the amount of links to the domain in the majority language and with the population that have the majority language as their first language.

2. *Literature review*

Bilingualism on the Web has been studied from various perspectives, but not yet extensively from a webometric point of view. Nagata et al. [16] and Li et al. [17] have studied how the content on the web could be used for translation purposes. The design of multilingual Web sites has been studied by Tonella et al. [18] and Cunliffe et al. [19]. Luna et al. [20] studied how to design culture-specific content for an optimal navigation experience for a multicultural audience. Cunliffe [21] studied how to promote minority languages on bilingual Web sites. Craven [22] studied the use of meta tags in

different languages. A study by Bar-Ilan and Gutman [23] showed that search engines ignore special characteristics of non-English languages.

A webometric study about linguistic patterns of academic Web use in Western Europe showed international interlinking patterns in English throughout Europe and in Swedish in Scandinavia [24]. Greece was found to be linguistically isolated on the Web, and Norway and Switzerland were not, although they are not members of the EU. Language seems to play a role in linking, at least in the academic Web.

Search engines can give researchers free access to huge amounts of data about the Web, but there are some challenges. No single search engine can cover the whole Web and the overlap between different search engines is quite low [25, 26, 27, 28, 29], but some topics may be covered better than others [8]. Vaughan and Thelwall [30] found significant differences between how major search engines cover commercial Web sites in different countries. Sites from the US were much better covered by the search engines compared with the other three countries, China, Taiwan and Singapore, in the study. It was concluded that the bias is not intentional, but that it is "*a natural result of cumulative advantage effects of US sites on the Web*". This means that one can't be sure to get all relevant documents when using a single search engine.

Several scholars have shown that results from search engines can be unreliable and unstable and the results can vary even within short periods of time [31, 32, 33, 34]. Bar-Ilan [35] showed a great deal of fluctuation and instability in the results of major search engines. Mettrop and Nieuwenhuysen [36] suggest that some of the instability and fluctuations may be caused by conscious decisions by the search engines to remove some Web sites and Web pages from their databases because of their content. Some search engines also try to remove duplicates. Search engines probably prioritize speed of delivering the results. It is better to give the most relevant results quickly than to give all possible results and use more time. So in order to prioritize speed search engines may not search through all of their databases and indexes, which may result in fluctuations. The use of multiple databases and multiple indexes which are not always identical may also cause fluctuations in the results. To improve response time, search engines may truncate the results and give an estimate of the number of hits.

Exactly how search engines operate is not known. Search engines use different ranking algorithms [37] and these are well kept business secrets. Google [38] only reveals that their ranking algorithm is "*based on a number of factors designed to provide end-users with helpful, accurate search results*". Some studies have found search engine AltaVista to have become more reliable and stable [8, 9, 14, 11]. Thelwall [39] found quite consistent results from quantitative analyses on the results from the search engines Google, Live Search and Yahoo. These later studies suggest that search engines may be valuable tools for data collection in webometric research.

Search engines offer some advanced search features like queries in certain languages or file types. Some search engines can also be used to search or calculate links and this makes them a promising data source for webometric studies.

3.   Background

Finland is a bilingual country with two official languages: Finnish and Swedish. Finland was part of Sweden until 1809 and this bond between the two countries is still visible through the minority languages in both countries. About 6% of the population in Finland speaks Swedish as their first language. Some municipalities, especially along the long coastline, are almost 100% Swedish-speaking. Language plays an important role even on the web. Some municipalities, with strong Swedish-speaking population, have their Web sites or parts of the Web sites in both languages. These municipalities usually have two domains, one in Finnish and one in Swedish, corresponding to the name of the municipality in these languages. The municipality of Pargas for instance, is Pargas in Swedish and Parainen in Finnish, and the municipality has two domains, www.pargas.fi and www.parainen.fi.

The region of Finland Proper is located in the South-Western parts of Finland and at the time of data collection in 2007 it had 54 municipalities [45]. There are five municipalities in the region that have Swedish as the majority language. These municipalities are Pargas (54% Swedish), Kimito (64.5%

Swedish), Nagu (71.4% Swedish), Houtskär (87.5% Swedish) and Korpo (74.1% Swedish)[1]. These municipalities are the only ones in the region that have two domains and at least parts of their Web sites in both Finnish and Swedish. These are also the municipalities studied in this research.

4.   Research questions

The goal of this research is to study whether there is a correlation between the majority language in the municipalities and inlinks. Does language play a part in linking patterns? This study will assess how municipal Web sites that have two domains in two different languages are linked to and whether there is a correlation between the proportion of inlinks to a domain in one language and the proportion of population that speak that language as their first language. This research will answer the following research question:

1. *When the majority of the population in a municipality has Swedish as their first language, does the Swedish domain of the municipality's Web site receive the majority of the inlinks?*

This research will also discuss any possible variations discovered between performances of the different search engines used for data collection.

5.   Methods

The five bilingual municipalities from the region of Finland Proper were selected for this study. Two types of inlinks to the investigated Web sites were collected: 1) inlinks from other municipalities in the region were collected with a Web crawler specially designed for information science purposes [40] and 2) inlink counts from the whole Web were collected with search engines Google, AltaVista and Yahoo. Thelwall [39] recommends Google for hit count estimates and Yahoo for other webometric purposes. AltaVista actually uses Yahoo's database [41] and it was included in the analysis mainly to

---

[1] The municipalities of Pargas, Nagu, Houtskär and Korpo merged on January 1, 2009, into the new municipality of Väståboland, and the municipality of Kimito merged together with two other municipalities into the new municipality of Kimitoön.

compare its results against those from Yahoo. The Web crawler was also used to determine the size of the Web sites in number of pages.

Although Google is at the moment the most popular search engine [42, 34, 43], it does not have as extensive advanced search features as some of the other search engines have. Therefore other search engines were also used for data collection. The *link:URL* command retrieves all pages linking to the given URL (in this study the homepages of the municipalities). This feature is available in Google, AltaVista and Yahoo! (through Yahoo's Site Explorer). AltaVista and Yahoo also provide a more extensive feature which can be used to retrieve all pages that have a link to any page inside the researched domains. Smith [44] argued that external inlinks are the most indicative ones because internal links are often navigational links; hence all internal links were excluded from the link counts. This was done with embedded advanced functions in both the Web crawler and the search engines. In Yahoo Site Explorer for instance, one can choose to show all inlinks "*except from this domain*" to exclude internal links. This procedure is however not available in Google and hence the results from Google may include navigational or self-linking links.

The proportion of the links to the domains in the majority language were compared with the proportion of population that have the majority language, Swedish, as their first language. Pearson correlation coefficients were calculated. Pearson correlation was also used to investigate whether there were a significant correlation between the size of the sites and the total amount of inlinks the municipalities receive.

6.   Results

Inlinks from other municipalities in the region It was studied to which of the two domains other municipalities in the region link to. Table 1 shows the proportion of links that target the Swedish domain of the municipalities and the proportion of population that speak Swedish as their first language. The link data was collected with a Web crawler [40]. The municipality of Houtskär received links from 6 other municipalities in the region. All of these link to the Swedish domain. Of the 12

municipalities linking to Kimito municipality, eight link to the Swedish domain (66.6%) and four link to the Finnish domain (33.4%). The municipality of Korpo was linked to by 9 municipalities, 8 of which (88.8%) link to the Swedish domain and only one (11.2%) to the Finnish domain. Nagu municipality's Swedish domain receives links from 7 (77.7%) other municipalities in the region and the Finnish domain is linked to by two municipalities (22.3%). The Swedish domain of municipality of Pargas, *pargas.fi*, is linked to by five other municipalities (55.5%), while the Finnish domain *parainen.fi* is linked to by four municipalities (44.5%). The Pearson correlation between the proportion of links to the Swedish domain and the proportion of population speaking Swedish as their first language is a very high 0.99 (significant at 5%, $p < 0,0001$). The majority of the links go to the domain in the majority language and they do so in almost same proportion as the majority language is spoken.

Table 1. Proportion of municipalities linking to the Swedish domains and
proportion of the population that speak Swedish as their first language

	Houtskär	Korpo	Nagu	Kimito	Pargas
Municipalities linking to the Swedish domain	100.0	88.8	77.7	66.6	55.5
First language Swedish	87.5	74.1	71.4	64.5	54.0

If we look at the actual number of links the researched municipalities receive from other municipalities in the region, the same tendency is visible. All domains in the majority language receive the majority of the interlinking links from other municipalities in the region (Table 2). All inlinks to Houtskär and almost 95% of the links to Korpo and Kimito are to the domain in the majority language. The Swedish domain of Pargas receives about 76% of the inlinks and the Swedish domain of Nagu receives 56% of the inlinks. The Pearson correlation between the proportion of inlinks and the proportion of the majority language in the municipalities is 0.88 (significant at 5%).

Table 2. Proportion of total inlinks to Swedish domains from municipalities in the region and

proportion of the population that speak Swedish as their first language

	Houtskär	Korpo	Nagu	Kimito	Pargas
Total interlinking links	100.0	94.8	56.0	93.7	76.3
First language Swedish	87.5	74.1	71.4	64.5	54.0

The strong correlations above suggest that the other municipalities in the region are well aware of the majority and minority languages in the researched municipalities and that this knowledge is visible in their linking.

Inlinks from the whole Web The link data collected with search engines was divided into two sets according to what kind of data was collected. The first analysis includes inlinks to the homepages of the researched municipalities (Figure 1) and the second set contains inlinks to any page in the Web sites (Figure 2).

The inlinks to the homepages were counted with AltaVista, Google and Yahoo. The proportions of inlinks to the Swedish domains of total inlinks to the municipalities Web sites can be seen in Figure 1 below. The first column in each case shows the results from AltaVista, the second column shows the results from Google and the third column shows the results from Yahoo. The fourth column for each municipality shows the proportion of the population that speak Swedish as their first language. The differences between the proportion of links collected with Yahoo and the proportion of the population speaking Swedish as their first languages are between 0.3% and 11.4%. In the case of Google the differences are between 4.6% and 30.4% and for AltaVista the differences are between 3.4% and 40.4%. If Korpo was excluded from the calculation, then the differences in the results from AltaVista would be between 3.4% and 6.7%. In the case of Pargas, although Swedish is the majority language in the municipality, the majority of the links target the domain in the minority domain. According to Google, majority of links to Nagu target the domain in the minority language and for Korpo the link

counts are even between the two domains. For the municipality of Korpo, AltaVista shows that only 33.7% of the links go to the domain in the majority language.
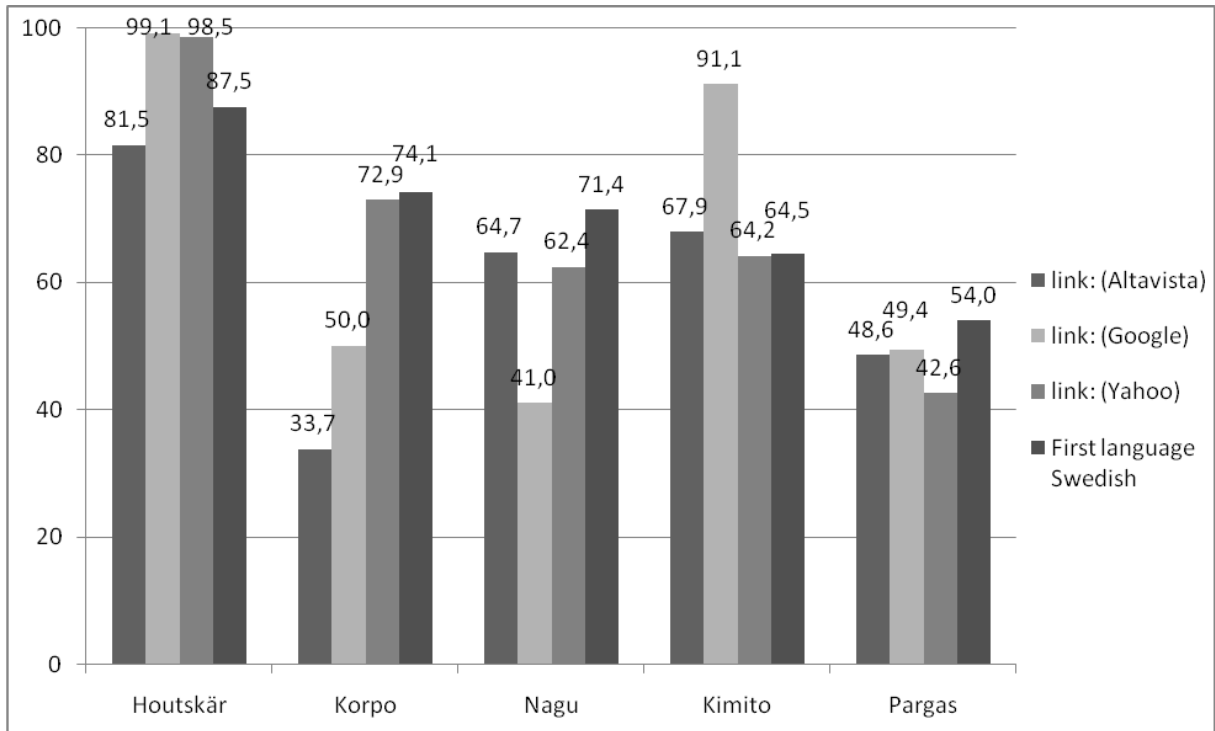

Figure 1. Proportion of links to homepages in Swedish and proportion of population with Swedish as the first language

The Pearson correlations between the proportion links and the proportion of population speaking Swedish as their first languages can be seen in Table 3 below. There is a strong correlation (Pearson 0.949, significant at 5%) between the results from Yahoo and language. The correlation between Google results and language is 0.671 (significant at 5%) and the correlation between AltaVista results and language is 0.620 (not significant at 5%). Although AltaVista and Yahoo use the same database, the correlation between their link counts is 0.66 (significant at 5%).

Table 3. Pearson correlations between proportion of links to the homepages in Swedish and proportion of population with Swedish as the first language

	AltaVista	Google	Yahoo	Language
AltaVista	-	**0.785**	**0.660**	0.620
Google	**0.785**	-	**0.771**	**0.671**
Yahoo	**0.660**	**0.771**	-	**0.949**
Language	0.620	**0.671**	**0.949**	-

The inlinks from the whole Web to any page in the target Web sites and excluding navigational and self-linking links were collected only with Yahoo and Altavista because this feature is not available in Google. When looking at the inlinks to the whole sites, the differences between the link counts calculated with Yahoo and the language are between 2.0% and 20.0% and with AltaVista between 0.5% and 29.9%. For Pargas again, the majority of links target the domain in the minority language.
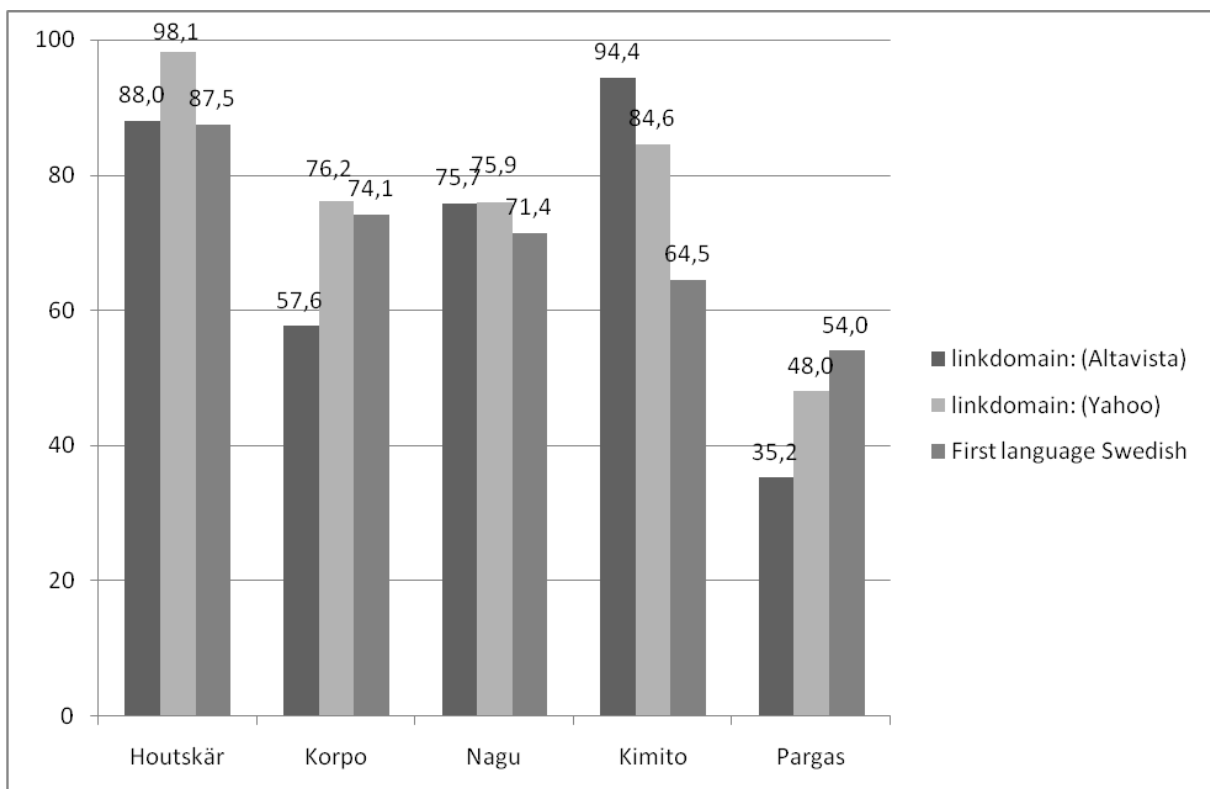

Figure 2. Proportion of links to whole sites and proportion of population with Swedish as the first language

The Pearson correlation coefficient for the correlation between the population that speak Swedish as their first language and the proportion of links collected with Yahoo to the Swedish domain is very high, almost 0.97 (significant at 5%). For the proportion of links collected with AltaVista and the proportion of the majority language, the correlation is also high, 0.81 (also significant at 5%). The correlation between the link data retrieved from Yahoo and the link data retrieved from AltaVista show a correlation of 0.925, which is a significant improvement when compared to the correlation of 0.66 when link counts only to the homepages were used. The correlations are visible in Table 4 below.

Table 4. Pearson correlation between proportion of links to the whole sites and
population with Swedish as the first language

	AltaVista	Yahoo	Language
AltaVista	-	**0.925**	**0.805**
Yahoo	**0.925**	-	**0.966**
Language	**0.805**	**0.966**	-

Significant correlations between how the Swedish domain receives the majority of links in those municipalities that the majority of population has Swedish as their first language were shown, and with that, the research question has been answered.

Correlation between Web site size and inlinks In contrast to earlier research there were no significant positive correlations between site size and inlink counts. The only significant correlation was a strong (Pearson -0.918, significant at 5%) negative correlation between site size and the total amount of inlinks retrieved with Yahoo's Site Explorer. In this study smaller Web sites attract proportionally more inlinks than larger ones. The site sizes and total inlink counts (to both domains combined) collected from different search engines are listed in Table 5 below.

Table 5. Number of pages and inlinks collected from different search engines and with different commands

	Pages	Google (to homepage)	Yahoo (to homepage)	Yahoo (to whole site)	AltaVista (to homepage)	AltaVista (to whole site)
Houtskär	427	107	1090	1240	536	927
Korpo	210	212	989	1221	679	1244
Nagu	81	39	1154	2003	1099	1862
Kimito	8862	45	346	1290	268	3496
Pargas	10590	85	594	1061	692	1502

7. Discussion

This research studied how bilingual Web sites with two domains in two languages attract links. The research question of this study was: *"When the majority of the population in a municipality have Swedish as their first language, does the Swedish domain receive the majority of the inlinks?"* Different methods to count links to the researched Web sites were used and strong correlations between the proportion of inlinks to the domain in Swedish and the proportion of population that have Swedish as their first language in the researched municipalities were shown, answering the research question. In the case of Pargas however, the majority of links did not target the domain in majority language with every used measure. This may be due to the fact that Pargas had the lowest Swedish majority, 54%. It is possible that the tendency to link to the majority language is clearer when there is a clearer difference between the majority and minority languages.

It is reasonable to assume that other municipalities in the same region, in some cases even neighboring municipalities, have the best knowledge about the language aspects of the researched municipalities. In that case it can also be assumed that the link counts from other municipalities give us a best case of how the languages affect linking and that therefore their link counts could be used as

reference when checking link data collected from search engines. We showed a strong positive correlation (Pearson, 0.88) between link counts and the languages in the municipalities. Therefore the search engine best suited for this kind of tasks should also give a strong positive correlation between the researched attributes. Of the used search engines, results from Yahoo were the most consistent and had the highest correlation with the proportion of languages, suggesting that Yahoo would be the best of the used search engines for webometric purposes, at least when language is a factor.

It is somewhat surprising that the correlation between link counts from Yahoo and AltaVista are as low as 0.66 when using links to the homepages. Although significant it is not even as good as the correlation between Yahoo and Google counts. This is surprising because AltaVista uses Yahoo's database and therefore should have access to the same data. This suggests the use of different retrieval algorithms and variations in the ways that the results are presented. What is also surprising is that in contrast to earlier research, there were no positive correlation between the size of the sites and inlink counts. In fact, there was a significant negative correlation suggesting that smaller sites receive proportionally more inlinks than larger sites. This could be due to the fact that some of the researched sites had a lot or even all of their content in both languages, increasing the site sizes significantly. It is possible that the results would have been different if the site sizes would have been measured and the link counts tested for the domains in both languages separately. This was not within the scope of this research, but would make an important subject for future studies.

Although the correlations are very strong and significant, there are some weaknesses with the present study. The use of search engines to calculate links may be considered as a weakness because it is still unclear exactly how different search engines present their results and whether they are withholding or excluding some of the results. The biggest weakness is the low number of cases researched. Five municipalities are too few to give results that could be generalized to every bilingual municipality in Finland. However the results are very clear and based on them we can say that there are clear indications for people linking to bilingual municipal Web sites to be well aware of the language aspects in the municipalities and for this knowledge to reflect on their linking behavior.

References

1. Vreeland, R.C. (2000). Law libraries in hyperspace: A citation analysis of World Wide Web sites. *Law Library Journal*, vol. 92, no. 1.

2. Chu, H., He, S. & Thelwall, M. (2002). Library and information science schools in Canada and USA: A webometric perspective. *Journal of Education for Library and Information Science*, vol. 43, no. 2.

3. Ingwersen, P. (1998). The calculation of Web impact factors. *Journal of Documentation*, vol. 54, no. 2, pp. 236-243.

4. Smith, A. & Thelwall, M. (2002). Web impact factors for Australasian Universities. *Scientometrics*, vol. 54, no. 1-2, pp. 363-380.

5. Thelwall, M. & Price, E. (2003). Disciplinary differences in academic web presence - A statistical study of the UK. *Libri*, vol. 53, no. 4, pp. 242-253.

6. Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. In the *Proceedings of the seventh international conference on World Wide Web 7*, pp.107-117, April 1998, Brisbane, Australia

7. Wilkinson, D., Harries, G., Thelwall, M. & Price, L. (2003). Motivations for academic web site interlinking: evidence for the Web as a novel source of information on scholarly communication. *Journal of Information Science*, vol. 29, no. 1, pp. 49-56.

8. Thelwall, M. (2001a). Extracting macroscopic information from Web links. *Journal of the American Society for Information Science and Technology*, vol. 52, no. 13, pp. 1157-1168.

9. Thelwall, M. (2002a). A comparison of sources of links for academic Web impact factor calculations. *Journal of Documentation*, vol. 58, no. 1, pp. 60-72.

10. Tang, R. & Thelwall, M. (2003). U.S. academic departmental Web-site interlinking in the United States Disciplinary differences. *Library & Information Science Research*, vol. 25, pp. 437-458.

11. Li, X., Thelwall, M., Musgrove, P. & Wilkinson, D. (2003). The relationship between the WIFs or inlinks of Computer Science Departments in UK and their RAE ratings or research productivities in 2001. *Scientometrics*, vol. 57, no. 2, pp. 239-255.

12. Thelwall, M. (2002b). Evidence for the existence of geographic trends in university Web site interlinking. *Journal of Documentation*, vol. 58, no. 5, pp. 563-574.

13. Holmberg, K. & Thelwall, M. (to appear, 2009). Local government web sites in Finland: A geographic and webometric analysis, *Scientometrics*.

14. Vaughan, L. & Thelwall, M. (2003). Scholarly use of the Web: What are the key inducers of links to journal Web sites? *Journal of the American Society for Information Science and Technology*, vol. 54, pp. 29-38.

15. Barjak, F. & Thelwall, M. (2008). A statistical analysis of the Web presences of European life sciences research teams. *Journal of the American Society for Information Science and Technology*, vol. 59, no. 4, pp. 628-643.

16. Nagata, M., Saito, T. & Suzuki, K. (2001). Using the Web as a bilingual dictionary. In the *Proceedings of the workshop on Data-driven methods in machine translation*, vol. 14, pp. 1-8.

17. Li, H., Cao, Y. & Li, C. (2003). Using bilingual Web data to mine and rank translations. *Intelligent systems*, vol. 18, issue 4, pp. 54-59.

18. Tonella, P., Ricca, F., Pianta, E. & Girardi, C. (2002). Restructuring multilingual web sites. In *Proceedings on Software Maintenance*, 2002, pp. 290-299.

19. Cunliffe, D., Jones, H., Jarvis, M., Egan, K., Huws, R. & Munro, S. (2002). Information architecture for bilingual Web sites. *Journal of the American Society for Information Science and Technology*, vol. 53, no. 10, pp. 866-873.

20. Luna, D., Peracchio, L.A. & Juan, M.D.de (2002). Cross-cultural and cognitive aspects of Web site navigation. *Journal of the Academy of Marketing Science*, vol. 30, no. 4, pp. 397-410.

21. Cunliffe, D. (2004). Promoting minority language use on bilingual Web sites. *Mercator Media Forum*, 2004.

22. Craven, T.C. (2002). Variations in use of meta tag descriptions by Web pages in different languages. *Information Processing & Management*, vol. 40, pp. 479-493.

23. Bar-Ilan, J. & Gutman, T. (2005). How do search engines respond to some non-English queries? *Journal of Information Science*, vol. 31, no. 1, pp. 13-28.

24. Thelwall, M., Tang, R. & Price, L. (2003). Linguistic patterns of academic Web use in Western Europe. Scientometrics, vol. 56, no. 3, pp. 417-432.

25. Lawrence, S. & Giles, C.L. (1998). Searching the World Wide Web. *Science*, vol. 280, pp. 98-100.

26. Lawrence, S. & Giles, C.L. (1999). Searching the Web: General and scientific information access. *IEEE Communications*, vol. 37, no. 1, pp. 116-122.

27. Thelwall, M. (2000). Web impact factors and search engine coverage. *Journal of Documentation*, vol. 56, no. 2, pp. 185-189.

28. Smith, A.G. (2003). Think local, search global? Comparing search engines for searching geographically specific information. *Online Information Review*, vol. 27, no. 2, pp. 102-109.

29. Jacsó, P. (2005). Visualizing overlap and rank differences among web-wide search engines. *Online Information Review*, vol. 29, no. 5, pp. 554-560.

30. Vaughan, L. & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes. *Information Processing & Management*, vol. 40, pp. 693-707.

31. Bar-Ilan, J. (1999). Search engine results over time – a case study on search engine stability. Cybermetrics, vol. 2/3, issue 1, paper 1. Retrieved February 28, 2007 from http://www.cindoc.csic.es/cybermetrics/articles/v2i1p1.pdf.

32. Rousseau, R. (1999). Daily time series of common single word searches in AltaVista and NorthernLight. *Cybermetrics*, vol. 2/3, no. 1, paper 2.

33. Snyder, H. & Rosenbaum, H. (2000). Can search engines be used as tools for Web-link analysis? A critical view. *Journal of Documentation*, vol. 55, no. 4, pp. 375-384.

34. Notess, G. (2005). Search Engine Statistics. *Search Engine Showdown*. Available at http://www.searchengineshowdown.com/statistics/ retrieved on March 4, 2008.

35. Bar-Ilan, J. (2000). Evaluating the stability of the search tools Hotbot and Snap: a case study. *Online Information Review*, vol. 24, no. 6, pp. 439-449.

36. Mettrop, W. & Nieuwenhuysen, P. (2001). Internet search engines – fluctuations in document accessibility. *Journal of Documentation*, vol. 57, no. 5, pp. 623-651.

37. Bar-Ilan, J. (2004). Comparing ranking of search results on the Web. *Information Processing & Management*, vol. 41, pp. 1511-1519.

38. Google (2004). Information for webmasters. *Google*. Retrieved from http://www.google.com/support/webmasters/bin/answer.py?answer=34432&topic=8524 on 25.2.2008.

39. Thelwall, M. (2008). Quantitative comparisons of search engine results, *Journal of the American Society for Information Science and Technology*, vol. 59, no. 11, pp. 1702-1710.

40. Thelwall, M. (2001b). A web crawler design for data mining. *Journal of Information Science*, vol. 27, no. 5, pp. 347-359.

41. SearchEngineWatch (2007). Major search engines and directories. Available at

http://searchenginewatch.com/2156221 retrieved on January 17, 2009.

42. Sullivan, D. (2005). Search engine sizes. *SearchEngineWatch*. Available at

http://searchenginewatch.com/showPage.html?page=2156481 retrieved on March 4, 2008.

43. Compete (2009). December Search Market Share: Holiday Bargain Hunting Supercharges Search,

Closes a Year of Google Gains. *Compete press release*, January 16, 2009. Retrieved from

http://blog.compete.com/2009/01/16/december-search-market-share-google-yahoo-live-ask-aol/ on

January 18, 2009.

44. Smith, A. (1999). A tale of two Web spaces: Comparing sites using Web impact factors. *Journal of*

Documentation, vol. 55, no. 5, pp. 577-592.

45. Local Finland (2009). Statistics. Available at

http://www.kunnat.net/k_peruslistasivu.asp?path=1;161;279;280;43252 retrieved on January 16, 2009.