

Brazil academic webuniverse revisited: A cybermetric analysis

ISIDRO F. AGUILLO¹

JOSÉ L. ORTEGA¹

BEGOÑA GRANADINO¹

1 CINDOC-CSIC. Joaquín Costa, 22. 28002 Madrid. Spain

{isidro;jortega;bgranadino}@cindoc.csic.es

Abstract

The analysis of the web presence of the universities by means of cybermetric indicators has been shown as a relevant tool for evaluation purposes. The developing countries in Latin-American are making a great effort for publishing electronically their academic and scientific results. Previous studies done in 2001 about Brazilian Universities pointed out that these institutions were leaders for both the region and the Portuguese speaking community. A new data compilation done in the first months of 2006 identify an exponential increase of the size of the universities' web domains. There is also an important increase in the web visibility of these academic sites, but clearly of much lower magnitude. After five years Brazil is still leader but it is not closing the digital divide with developed world universities.

Keywords:

Cybermetrics, Brazilian universities, Web indicators, Websize, Web visibility

1. Introduction.

The Internet Lab (CINDOC-CSIC) has been working on the development of web indicators for the Latin-American academic sector since 1999 (1,2,3,4,5,6,7,8). The region is especially interesting due to the presence of several very large developing countries that are perfect targets for case studies and the use of Spanish and Portuguese languages in opposition to English usually considered the scientific "lingua franca".

During the Fifth RICYT conference, a webometric analysis of Brazilian universities was presented (2). This contribution reviews the current status (January 2006) of the web presence of the Brazilian higher education institutions providing a comparative analysis after five years. The web indicators used are basically the same (size, visibility, popularity) as well the methodology to compile them with search engines. Unfortunately most of the databases are no longer the same and direct comparison is not possible. We opted for using maximum values of combined results of the several engines available during the two sampling periods.

2. Methods.

A Catalogue with comprehensive lists of URL addresses is maintained by the authors as part of the Webometrics Ranking of World Universities (www.webometrics.info), where identification and

compilation of institutions criteria and search engine data extraction methods are explained. It is relevant for the purposes of future scientometric analysis that only universities with independent institutional web domain are considered.

This can explain why the number of Brazilian universities studied in October 2001 (158 institutions) is lower than those identified during January 2006 (178 institutions), which now consider more small and private institutions than before. The data was recovered using the following search engines: Alltheweb, Altavista, Google, Northern Light and Hotbot (2001) and Google, MSN Search, Yahoo Search and Teoma (2006). The popularity was extracted for both periods from Alexa database.

The indicators calculated are size (number of pages), visibility (number of external inlinks) and popularity (number of visits intercepted by Alexa). In 2006 the number of rich files (documents in pdf, ps, doc or ppt formats) was obtained for each university.

The co-linking phenomena were analysed using data also collected during January 2006 from Yahoo search engine according to the following search strategy:

linkdomain:universityA +site:universityB

Excluding several really small universities with few inlinks, a total of 167 academic institutions of Brazil were used to carry out a co-link map (9). In order to built this map an asymmetrical matrix was used between the links of these 167 academic web sites.

The cosine (10,11) was used to calculate the similarity among the web sites.

$$Sim(S_i, S_j) = \frac{\sum e_i e_j}{\sqrt{\sum (e_i)^2 \sum (e_j)^2}}$$

The matrix was turn to distances and a MDS map ($\phi=0,079$) was carried out. The map was plotted with the PROXSCAL module of SPSS 13.0 (www.spss.com) and finally it was displayed in a network graph with UCINET 6.1 and NetDraw 2.2. (www.analytictech.com).

3. Results.

The results show that the number of web pages has increased exponentially. From Table 1 that reflects only the larger universities, it can be seen that the maximum value among search engines is more than 15-40 times bigger than five years ago. Although the engine coverage has expanded considerably in that period, the huge increase seems genuine. The ESCOLA PAULISTA OF MEDICINE case is not unique as there are several other universities with a low presence in the Web that in the last years has increased its volume of knowledge in the Web by several orders of magnitude.

Table 1: Top 20 Brazilian universities according to Websize

UNIVERSITY	DOMAIN	PAGES		TIMES LARGER
		2001	2006	
UNIVERSITY OF SAO PAULO	<i>usp.br</i>	121489	4520000	37
STATE UNIVERSITY OF CAMPINAS	<i>unicamp.br</i>	81455	3520000	43
FEDERAL UNIVERSITY OF SANTA CATARINA	<i>ufsc.br</i>	66686	1150000	17
FEDERAL UNIVERSITY OF MINAS GERAIS	<i>ufmg.br</i>	41200	872000	21
ESCOLA PAULISTA OF MEDICINE	<i>unifesp.br</i>	5560	862000	155
FEDERAL UNIVERSITY OF PARANA	<i>ufpr.br</i>	31700	833000	26
FEDERAL UNIVERSITY OF RIO DE JANEIRO	<i>ufrj.br</i>	43800	811000	19
FEDERAL UNIVERSITY OF RIO GRANDE DO SUL	<i>ufrgs.br</i>	47300	785000	17

FEDERAL UNIVERSITY OF BAHIA	<i>ufba.br</i>	18300	767000	42
PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO	<i>puc-rio.br</i>	29500	750000	25
UNIVERSITY OF BRASILIA	<i>unb.br</i>	28900	675000	23
STATE UNIVERSITY PAULISTA	<i>unesp.br</i>	44100	558000	13
FEDERAL UNIVERSITY OF PERNAMBUCO	<i>ufpe.br</i>	30100	345000	11
FEDERAL UNIVERSITY OF SANTA MARIA	<i>ufsm.br</i>	12300	320000	26
PONTIFICAL CATHOLIC UNIVERSITY OF PARANA	<i>pucpr.br</i>	13200	313000	24
UNIVERSITY OF CAXIAS DO SUL	<i>ucs.br</i>	1011	257000	254
PONTIFICAL CATHOLIC UNIVERSITY OF RIO GRANDE DO SUL	<i>pucrs.br</i>	22200	242000	11
STATE UNIVERSITY OF RIO DE JANEIRO	<i>uerj.br</i>	11000	234000	21
PONTIFICAL CATHOLIC UNIVERSITY OF SAO PAULO	<i>pucsp.br</i>	7550	217000	29
FEDERAL UNIVERSITY FLUMINENSE	<i>uff.br</i>	10500	167000	16

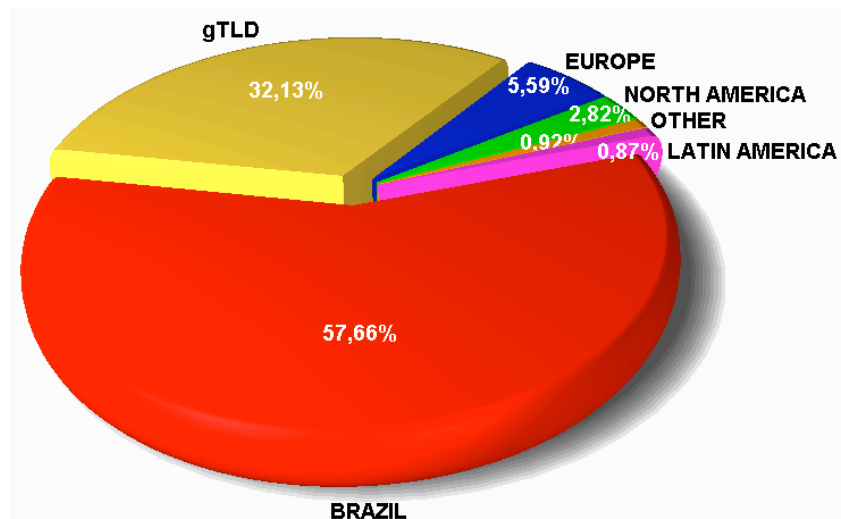
Visibility results (Table 2) were obtained from the number of external inlinks received by each institutional domain (Altavista in 2001; Yahoo and MSN in 2006). Data show that the number of inlinks has also been multiplied by a factor of 4 to 10 (even more when considering small universities). This is clearly a remarkable achievement but the ratio is far lower than the observed for the website.

Table 2: Top 20 Brazilian universities according to Web visibility

UNIVERSITY	DOMAIN	INLINKS		TIMES LARGER
		2001	2006	
UNIVERSITY OF SAO PAULO	usp.br	46481	367985	8
STATE UNIVERSITY OF CAMPINAS	unicamp.br	41697	157374	4
PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO	puc-rio.br	9750	93323	10
FEDERAL UNIVERSITY OF RIO DE JANEIRO	ufrj.br	15629	75061	5
FEDERAL UNIVERSITY OF SANTA CATARINA	ufsc.br	12995	74463	6
FEDERAL UNIVERSITY OF RIO GRANDE DO SUL	ufrgs.br	14855	72479	5
FEDERAL UNIVERSITY OF MINAS GERAIS	ufmg.br	10187	41900	4
UNIVERSITY OF BRASILIA	unb.br	7320	41139	6
STATE UNIVERSITY PAULISTA	unesp.br	6504	27700	4
FEDERAL UNIVERSITY OF PARANA	ufpr.br	6979	27000	4
FEDERAL UNIVERSITY OF BAHIA	ufba.br	6212	26200	4
FEDERAL UNIVERSITY OF PERNAMBUCO	ufpe.br	6413	22072	3
STATE UNIVERSITY OF RIO DE JANEIRO	uerj.br	5015	21000	4
PONTIFICAL CATHOLIC UNIVERSITY OF RIO GRANDE DO SUL	pucrs.br	3825	17600	5
FEDERAL UNIVERSITY OF RIO GRANDE DO NORTE	ufrn.br	2272	16100	7
PONTIFICAL CATHOLIC UNIVERSITY OF SAO PAULO	pucsp.br	2936	15902	5
FEDERAL UNIVERSITY FLUMINENSE	uff.br	3701	15700	4
FEDERAL UNIVERSITY OF SANTA MARIA	ufsm.br	4216	11900	3
FEDERAL UNIVERSITY OF CEARA	ufc.br	689	11800	17
STATE UNIVERSITY OF PIAUI	uespi.br	122	11200	92

If we classified the inlinks according to the domains from they were originated, most of them are from Brazil itself (Figure 1). It is not possible to know the exact link origin of the generic Top Level Domains (gTLD: com, org, net, int, info, biz) but they are responsible for most than $\frac{3}{4}$ of the rest of the inlinks. They are followed by North American domains (mostly US universities under edu domain, plus .us, .gov., .mil and .ca domains) and the European domains, both with larger contributions than the Latin-American countries.

Figure 1. Inlinks distribution by region of origin



Global impact was measured using the Webometrics Rank (WR). The relative positions of the top universities have not changed between the two periods (Table 3) and Brazil is leader among the Latin-American universities. But globally the position has been eroded when compared with EU and specially US universities.

Table 3: Webometrics Rank for the Top Brazilian Universities

UNIVERSITY	DOMAIN	WR (BR) (WORLD)		
		2001	2006	
UNIVERSITY OF SAO PAULO	<i>usp.br</i>	1	1	117
STATE UNIVERSITY OF CAMPINAS	<i>unicamp.br</i>	2	2	254
FEDERAL UNIVERSITY OF SANTA CATARINA	<i>ufsc.br</i>	5	3	505
FEDERAL UNIVERSITY OF RIO DE JANEIRO	<i>ufrj.br</i>	3	4	527
FEDERAL UNIVERSITY OF RIO GRANDE DO SUL	<i>ufrgs.br</i>	4	5	597
PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO	<i>puc-rio.br</i>	7	6	654
UNIVERSITY OF BRASILIA	<i>unb.br</i>	8	7	734
FEDERAL UNIVERSITY OF MINAS GERAIS	<i>ufmg.br</i>	6	8	768
STATE UNIVERSITY PAULISTA	<i>unesp.br</i>	9	9	848
FEDERAL UNIVERSITY OF PARANA	<i>ufpr.br</i>	11	10	912

The popularity report, measured by the number of visits intercepted by Alexa and then ranked against a worldwide domain database, shows similar results (Table 4), although University of Sao Paulo is now among the 10,000 most visited websites of the world.

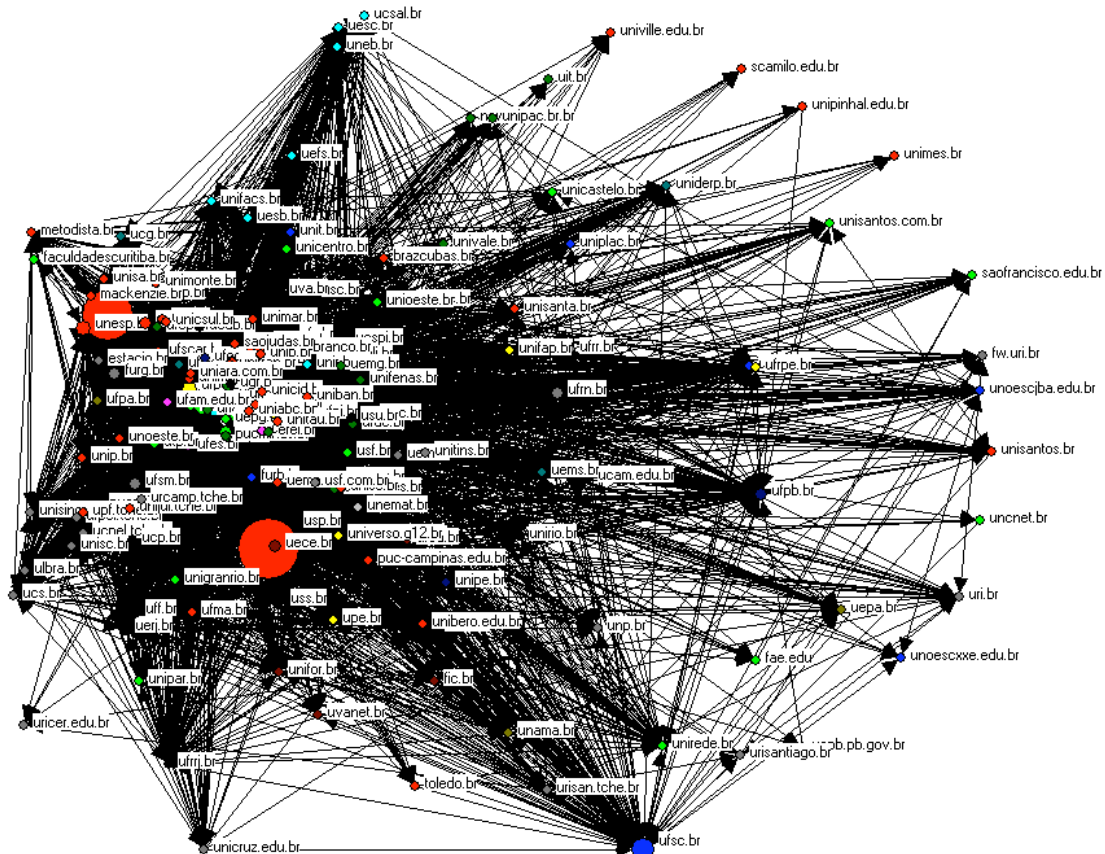
Table 4: Top 20 Brazilian universities according to Web popularity

UNIVERSITY	DOMAIN	ALEXA RANKING	
		2001	2006
UNIVERSITY OF SAO PAULO	<i>usp.br</i>	30982	7019
FEDERAL UNIVERSITY OF RIO DE JANEIRO	<i>ufrj.br</i>	140136	14015
STATE UNIVERSITY OF CAMPINAS	<i>unicamp.br</i>	58910	20895
UNIVERSITY OF BRASILIA	<i>unb.br</i>	64989	22451
FEDERAL UNIVERSITY OF RIO GRANDE DO SUL	<i>ufrgs.br</i>	2846	26063
FEDERAL UNIVERSITY OF SANTA CATARINA	<i>ufsc.br</i>	171395	30560
FEDERAL UNIVERSITY OF MINAS GERAIS	<i>ufmg.br</i>	140135	32997

PONTIFICAL CATHOLIC UNIVERSITY OF RIO DE JANEIRO	<i>puc-rio.br</i>	83428	33863
UNIVERSITY ESTACIO DE SA	<i>estacio.br</i>	33304	42506
UNIVERSITY NORTE DO PARANA	<i>unopar.br</i>	392651	50086
FEDERAL UNIVERSITY OF PARANA	<i>ufpr.br</i>	207920	51881
FEDERAL UNIVERSITY OF BAHIA	<i>ufba.br</i>	170865	51962
PONTIFICAL CATHOLIC UNIVERSITY OF MINAS GERAIS	<i>pucminas.br</i>	95945	52925
STATE UNIVERSITY PAULISTA	<i>unesp.br</i>	152536	57217
ESCOLA PAULISTA OF MEDICINE	<i>unifesp.br</i>		65446
FEDERAL UNIVERSITY OF PARA	<i>ufpa.br</i>	167641	65710
UNIVERSITY PAULISTA	<i>unip.br</i>	40495	68049
PONTIFICAL CATHOLIC UNIVERSITY OF SAO PAULO	<i>pucsp.br</i>	164017	72590
UNIVERSITY PRESBITERIANA MACKENZIE	<i>mackenzie.br</i>	79470	79294
STATE UNIVERSITY OF RIO DE JANEIRO	<i>uerj.br</i>	270067	81035

The Figure 2 shows the co-link map among 167 Brazilian academic institutions. The size of the nodes shows the number of pages of each institution and the colour is the State that each university.

Figure 2. Co-link Map of Brazilian Universities



The map shows that the bigger universities are the University of Sao Paulo (*usp.br*) and The University State of Campinas (*unicamp.br*) both in the Sao Paulo State. It can see some regional pattern with the universities of Bahia Estate (sky blue) and the universities of Rio Grande do Sul State (grey) and Ceará State (maroon). The Sao Paulo State is the state with more universities in Brazil. It is also appreciate that the federal universities are located in the centre and the state universities are more peripheral. This allows us to think that the federal universities are the main universities in the Brazilian academic web network.

4. Conclusions.

The commitment of the Brazilian universities to the Web has increased notably in the last five years, so many of the web indicators have jumped to levels several times higher than before. However, the global webometrics rank of these institutions is still far from those of their developed countries' counterparts.

The gap of the digital divide is widening, being a subject of major concern. Brazil should increase at faster rates its web visibility, publishing more electronic documents through institutional or subject repositories, offering more added value information in English and reinforcing cooperation with other institutions of the Latin-American region and the rest of the World.

References

1. González Martín, R.; Aguillo, I. F. La presencia de las Universidades iberoamericanas en Internet: Un estudio cibernético en el "cono sur".IV Taller Iberoamericano/Interamericano de Indicadores de Ciencia y Tecnología. México, DF, 12-14 July 1999. www.ricyt.edu.ar/interior/normalizacion/IV_taller/aguillo.pdf
2. Aguillo, I. F. Propuesta para la incorporación de indicadores cibernéticos al sistema de metadatos Dublin Core. V Taller Iberoamericano e Interamericano de Indicadores de Ciencia y Tecnología. Montevideo (Uruguay), 15-18 October 2001. <http://www.ricyt.edu.ar/interior/interior.asp?Nivel1=6&Nivel2=2&IdTaller=8&Idioma=>
3. Aguillo, I. F. Cibermetría con motores de búsqueda. Indicadores Web de Universidades Iberoamericanas. Tercer Taller Iberoamericano de Bibliometría. Madrid, Spain, 3-5 March 2003
4. Aguillo, I. F. Evolución reciente de los Indicadores cibernéticos de las Universidades Iberoamericanas. Congreso: II Seminario Internacional sobre Estudios Cuantitativos y Cualitativo de la Ciencia y la Tecnología "Prof. Gilberto Sotolongo Aguilar" INFO 2004. La Habana, Cuba, 15 April 2004
5. Aguillo, I. F. (2005). Indicadores de contenidos para la web académica iberoamericana. BiD: textos universitarios de biblioteconomía i documentació, desembre, núm. 15. http://www2.ub.edu/bid/consulta_articulos.php?fichero=15aguil2.htm
6. Aguillo, I. F.; Granadino, B. & Llamas, G. (2005). Posicionamiento en el Web del sector académico iberoamericano. *Interciencia*, 30(12):1-5
7. Aguillo, I. F.; Granadino, B.; Ortega, J.L. & Prieto, J.A. What the Internet says about Science. *The Scientist*, 2005, Volumen: 19(14): 10
8. Thelwall, M. & Aguillo, I. F. La salud de las web universitarias españolas. *Rev. Esp. Documentación Científica*, 2003, vol 26(3): 291-305
9. Ortega-Priego, J.L., & Aguillo, I.F. (2005) A Web map of the SCIC research centres: a comparative study of the cosine and Pearson's r. In P. Ingwersen & B. Larsen (Eds.). Stockholm: Karolinska University Press.
10. Salton, G. (1971). The Smart Retrieval System~Experiments. In G. Salton (Ed.), *Automatic Document Processing*. New York: Prentice-Hall, Englewood Cliffs.
11. Salton, G., Wong, A., & Yang, C.S. (1975). A Vector Space Model for automatic indexing. *Communications of the Association for Computing Machinery*, 18(11), 613-620.