

Indicadores Web de actividad científica formal e informal en Latinoamérica

Web indicators for describing formal and informal scholarly communication in Latin America

Isidro F. Aguillo*, José Luis Ortega*, José Antonio Prieto*,
Begoña Granadino*

Resumen. Utilizando los datos cibernéticos de las 500 primeras universidades latinoamericanas de acuerdo a indicadores Web obtenidos mediante motores de búsqueda, se analiza la comunicación de sus actividades científicas y académicas a través de Internet. Se presta especial atención al uso de los llamados ficheros ricos, normalmente asociados a documentos ligados a la publicación científica. Además, el uso de estos formatos para la comunicación informal se está generalizando entre las instituciones líderes de la región.

Palabras clave: cibermetría, indicadores Web, actividad científica, comunicación científica informal, universidades latinoamericanas.

Abstract. The top 500 Latin American universities according to Web indicators derived from search engines were selected in order to obtain webometric data for the analysis of their commitment to the publication of their scientific effort in the Net. The target of the study were the rich files, a format usually devoted to documents intended to be used as scientific or academic publications. Moreover, these file types are increasingly used by the Latin American leader institutions as a mean for informal scholarly communication.

Keywords: webometrics, web indicators, scientific effort, informal scholarly communication, Latin American universities.

1. Introducción

Los estudios que utilizan las publicaciones en la Web para derivar patrones cuantitativos de la actividad y producción científica son numerosos (Bjorneborn y Ingwersen, 2004), aunque intentan aplicar un marco teórico, metodologías y unidades ajenas a dicho entorno. La bibliometría tradicional puede aplicarse sin mayores problemas al

* Laboratorio de Cibernética. CINDOC-CSIC. Madrid.

Correo-e: {isidro;jortega;joseaprieto;bgranadino}@cindoc.csic.es.

Recibido: 14-12-06.

Una versión preliminar fue presentada al III Seminario Internacional sobre Estudios Cuantitativos y Cualitativos de la Ciencia y la Tecnología «Profesor Gilberto Sotolongo Aguilar». Congreso Internacional de Información-Info 2006. La Habana, Cuba, 17-21 abril del 2006.

estudio de las revistas científicas electrónicas y el análisis de citas bibliográficas entre publicaciones Web o con versiones en papel, no resulta doctrinalmente diferente. Las iniciativas que proponen contabilizar también el número de consultas o volcados de artículos en su edición electrónica (Harnad y Brody, 2004) no se alejan del modelo de evaluación tradicional. Sin embargo, la web ofrece contenidos mucho más amplios y variados que las publicaciones periódicas, organizados en unidades discretas distintas y además, las motivaciones detrás de los enlaces hipertextuales no se ciñen, ni mucho menos, a la indicación de una referencia bibliográfica (Thelwall, 2003).

La necesidad de reconocer nuevas unidades, de desarrollar indicadores más variados, de identificar patrones de enlaces más generales y de incorporar los aspectos informales a la evaluación de la actividad científica publicada a través de la web ya fue propuesta por Aguillo (2002). En dicho trabajo se proponía utilizar los llamados ficheros ricos de los subdominios académicos para enfatizar su uso diferencial por parte de dicha comunidad. Estos indicadores han sido incorporados a estudios empíricos (Aguillo et al., 2005b) que pretenden la descripción de escenarios globales a partir de datos cibernéticos.

El escenario latinoamericano ha sido objeto de estudio anteriormente por el Laboratorio de Cibermetría del CINDOC-CSIC (Aguillo, 2005; Aguillo et al., 2005a), aunque dichos trabajos estaban más centrados en el estudio de la visibilidad hipertextual.

2. Objetivos

El uso de indicadores Web es complementario al escenario ofrecido por los métodos bibliométricos (Wilkinson et al., 2003; Aguillo et al., 2006). El objetivo de este trabajo es ofrecer una serie de medidas que amplíen las posibilidades de evaluación de las actividades académicas y científicas, para incluir no solo las que conducen a la publicación de artículos en revistas de calidad, sino también los resultados de carácter más informal que representan procesos de comunicación científica realizados a través de Internet.

La comunicación informal se realiza a través de depósitos institucionales o temáticos, también llamados repositorios, y que pueden incluir tanto copias de trabajos ya publicados, como borradores en fase final, o material inédito (tesis y similares). Esta comunicación informal también se lleva a cabo a través de páginas personales mediante el auto-archivo del científico o grupo de investigación responsable de la sede (Swan, 2005). Estos comportamientos forman parte de las llamadas iniciativas de «Open Access» (<http://www.soros.org/openaccess/>), donde el acceso libre a la información científica se ve como una estrategia para incrementar significativamente su difusión a audiencias más amplias, especialmente aquellas situadas en los países en vías de desarrollo.

El presente estudio se centra en el análisis de las páginas Web que, por su formato, pueden representar alguna de las etapas de los procesos de comunicación científ-

ca, incluyendo las publicaciones formales tras revisión por pares, pero también *pre-prints* y documentos informales u orientados a la formación o divulgación científica.

3. Metodología

Se han utilizado como fuente de información los motores de búsqueda en la Web, habiéndose seleccionado por sus capacidades cibernéticas los 4 más grandes: Google (www.google.com), Yahoo! Search (search.yahoo.com); MSN Search (search.msn.com) y Teoma (www.teoma.com). De acuerdo con el indicador combinado (WR) propuesto en el *Webometrics Ranking of World Universities* (www.webometrics.info) se han seleccionado las primeras 500 universidades latinoamericanas (datos de enero de 2006).

El indicador WR ha sido descrito por Aguillo et al. (2006) que comprueban que genera clasificaciones tanto de productividad como de impacto similares a las de rankings basados en datos bibliométricos. Entre las ventajas que ofrece se incluye la disminución de los sesgos en contra de las instituciones de carácter tecnológico o de las grandes universidades de los países en vías desarrollo. En el aspecto negativo podemos indicar que sobrevalora las actividades no académicas y es muy sensible a las comunidades lingüísticas, aunque en este caso la región seleccionada presenta bastante homogeneidad en este sentido.

Para cada una de ellas se ha obtenido su posición absoluta y relativa, el tamaño expresado en número de páginas Web, la visibilidad medida a través del número de enlaces externos recibidos y el volumen de ficheros ricos según los datos recopilados por el buscador Google. La estrategia utilizada en este último caso es:

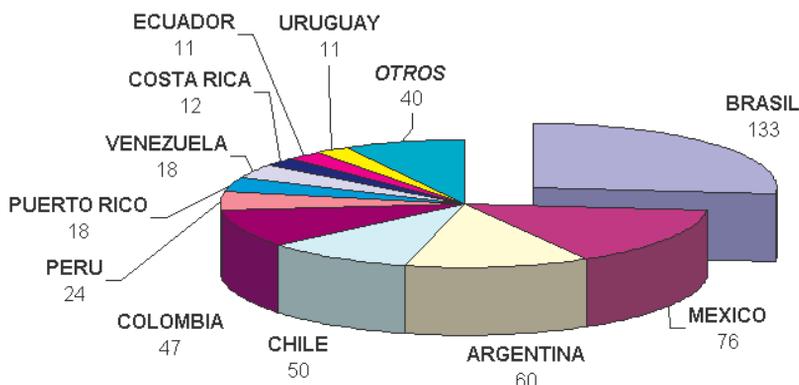
+site:dominio +filetype:fichero rico

donde fichero rico puede adoptar los valores *pdf* (Adobe Acrobat), *doc* (MS Word), *ps* (Postscript) o *ppt* (MS Powerpoint). Otros formatos fueron considerados y descartados por el bajo número de resultados obtenidos respecto a los anteriores. Las altas cifras obtenidas para los formatos pdf y doc exceden claramente del número requerido para aspectos administrativos o burocráticos y reflejan por tanto patrones de comunicación, similares a los más obvios de los formatos ps y ppt.

4. Resultados

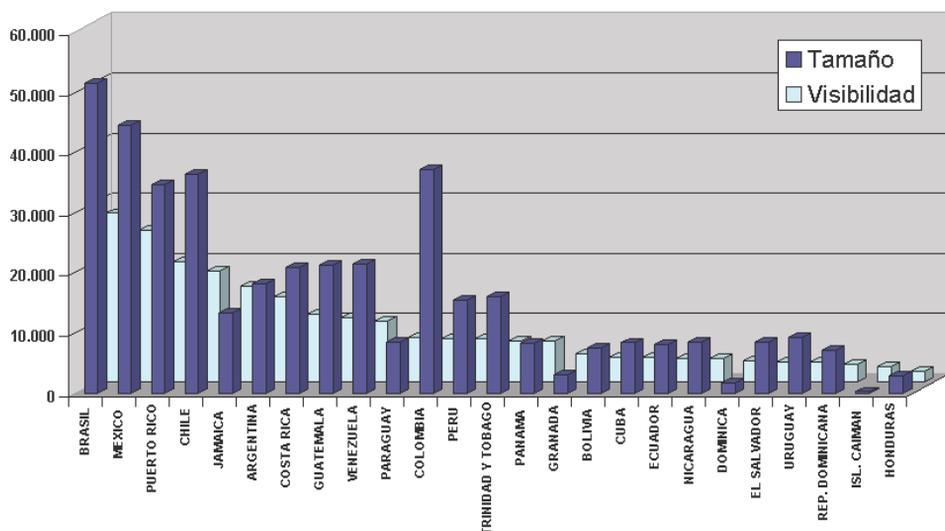
Un total de 25 países están representados en la muestra, siendo Brasil con 133 universidades y México con 76 los que tienen más instituciones entre las 500 primeras (figura 1). La región representa sólo el 2,0% del total de las 1.000 primeras universidades del mundo, aunque dicho porcentaje sube al 6.8% cuando consideramos las 5.000 primeras.

Figura 1
Distribución por países de las 500 primeras universidades latinoamericanas de acuerdo a indicadores Web (Webometrics Rank, enero 2006)



Las medidas del tamaño se obtuvieron de la combinación de datos de los cuatro motores, pero dada la irregularidad de su comportamiento (Bar-Ilan, 2005), se excluyeron los valores máximo y mínimo. La visibilidad se obtuvo a través de los motores Yahoo! Search y MSN Search, dada la imposibilidad de derivar esta medida desde Google.

Figura 2
Media de páginas (tamaño) y enlaces recibidos (visibilidad) por universidad en los países representados en la muestra (enero 2006)

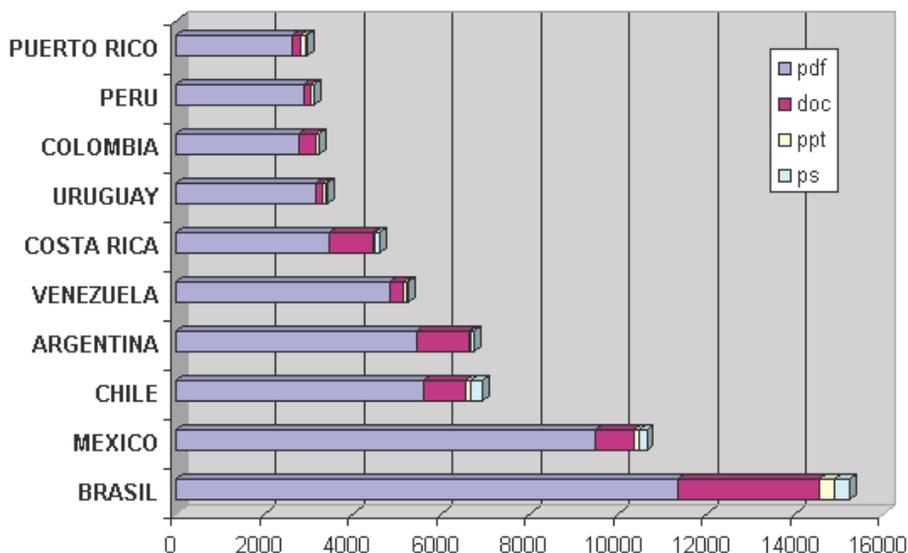


Tal como se muestra en la figura 2, Brasil no sólo es el país mejor representado en la muestra, sino que también es el que mayor tamaño medio por universidad y más enlaces externos recibe, clasificación esta última donde destacan los países que utilizan el inglés (Jamaica, Puerto Rico). Ello significa que las universidades brasileñas, además de numerosas entre las 500 primeras, ocupan en dicha lista posiciones de liderazgo. Hay que destacar el escaso volumen de páginas de las universidades argentinas respecto a lo esperado y la baja visibilidad de las colombianas, a pesar de su gran tamaño medio (¿motivado quizá por la indización de contenidos de la Internet invisible?). Habría que realizar un estudio al respecto, pero datos provisionales muestran un número muy bajo de páginas en inglés, lengua franca de la comunicación científica, en estas universidades.

La Universidad de Puerto Rico aparece con dos dominios diferentes en la lista, que no son idénticos y que por tanto requieren entradas distintas. Ello no afecta significativamente a la medida de tamaño, pero sí (negativamente) a la medida de la visibilidad que se resiente del reparto de enlaces entre dos dominios.

El estudio de la producción más ligada a procesos de comunicación tanto académica como científica se centra en el volumen de ficheros ricos y su reparto según tipología entre las universidades de la región.

Figura 3
Distribución de los ficheros ricos según formato y país
(media de universidades, enero 2006)



El formato Adobe Acrobat (pdf) es consistentemente el más utilizado (figura 3). La edición de este tipo de documentos requiere el manejo de un programa habitualmente no disponible en los paquetes ofimáticos convencionales (tales como MS Office). Aunque es posible convertirlo desde otros formatos, esta dificultad nos indica que se reserva para situaciones donde se valora su universalidad de facto como formato de comunicación de documentos por la Web.

Aunque un número significativo de ficheros pdf se pueden asociar a procesos administrativos o burocráticos (memorias internas, formularios), las cifras altas observadas en la tabla I (docenas o cientos de miles, es decir órdenes de magnitud por encima de su producción ISI) nos indican que un porcentaje muy elevado es en realidad utilizado en servicios de comunicación académica o científica: repositorios de artículos, bases de datos de tesis a texto completo o documentos (incluidos borradores) en páginas personales de profesores e investigadores.

Tabla I
Universidades con más ficheros pdf (enero 2006)

	<i>Doc. ACROBAT (pdf)</i>
UNIVERSIDAD DE SAO PAULO	307.000
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO	282.000
UNIVERSIDAD ESTATAL DE CAMPINAS	166.000
UNIVERSIDAD DE BRASILIA	147.000
UNIVERSIDAD DE GUADALAJARA	123.000
PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE	112.000
UNIVERSIDAD FEDERAL DE RIO DE JANEIRO	108.000
UNIVERSIDAD DE BUENOS AIRES	98.100
UNIVERSIDAD FEDERAL DE SANTA CATARINA	90.700
UNIVERSIDAD FEDERAL DE RIO GRANDE DO SUL	67.600

Un análisis detallado de las 10 universidades con mayor volumen de documentos en formato pdf muestra que dicha lista coincide con la de las universidades de mejor posición Web, lo que implica que una alta cantidad de publicaciones redundan positivamente en la visibilidad de las instituciones. Muchas de ellas son cabecera o centro referente de sus respectivos sistemas académicos, lo que sugiere que son pioneros en las iniciativas *Open Access* de cada país o que incluso centralizan esfuerzos cooperativos en este sentido.

El formato MS Word (doc) es extremadamente popular y su publicación en la Web es más fácil que en el caso anterior. Sin embargo, es mucho menos habitual (Figura 3), lo que sugiere su utilización como alternativa o complemento al pdf y posiblemente en este caso los documentos tengan un carácter más administrativo o, en el caso de publicaciones, de borrador previo.

Tabla II
Universidades con más ficheros doc (enero 2006)

	<i>DOC. MS WORD (doc)</i>
UNIVERSIDAD FEDERAL DE RIO DE JANEIRO	56.600
UNIVERSIDAD DE SAO PAULO	52.600
UNIVERSIDAD DE BUENOS AIRES	38.100
UNIVERSIDAD FEDERAL DE SANTA CATARINA	32.400
UNIVERSIDAD ESTATAL DE CAMPINAS	26.500
UNIVERSIDAD ESTATAL DE SANTA CATARINA	24.800
UNIVERSIDAD DE CHILE	24.500
UNIVERSIDAD FEDERAL DE PARANA	24.000
UNIVERSIDAD FEDERAL DE MINAS GERAIS	22.100
UNIVERSIDAD DE BRASILIA	20.500

Es significativa la ausencia en la tabla II de la UNAM, que podría explicarse por la existencia de políticas editoriales o un fuerte soporte informático institucionalizado que facilitarían el uso del Acrobat. Algunas subdominios de la UNAM sugieren esto último.

Si en los casos anteriores cabría dudar de la importancia del formato en la comunicación de la actividad investigadora, con respecto al Postscript dicha posibilidad se puede descartar. Se trata de un formato cuyo uso está muy centrado en las comunidades de ingenieros, matemáticos, físicos o cualquier otro que requiera el uso de formulas y notaciones matemáticas complejas. El ps resulta la opción elegida en esos colectivos para la publicación de sus resultados y, por tanto, la abundancia de ellos en una sede universitaria informa de su fuerte carácter tecnológico. La tabla III muestra nuevos actores con respecto a las anteriores, pero también la presencia de las grandes instituciones generalistas que albergan también escuelas y facultades tecnológicas muy importantes (Costa Rica, Colombia, Venezuela).

Tabla III
Universidades con más ficheros ps (enero 2006)

	<i>DOC. POSTSCRIPT (ps)</i>
UNIVERSIDAD DE SAO PAULO	24.600
UNIVERSIDAD DE CHILE	12.200
UNIVERSIDAD ESTATAL DE CAMPINAS	12.000
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO	10.600
UNIVERSIDAD NACIONAL DE CORDOBA	2.690
UNIVERSIDAD ESTATAL PAULISTA	2.600
UNIVERSIDAD DE COSTA RICA	1.070
UNIVERSIDAD DE BRASILIA	916
UNIVERSIDAD DE LOS ANDES BOGOTA	729
UNIVERSIDAD SIMON BOLIVAR VENEZUELA	673

Al igual que en el caso anterior, el uso del formato Powerpoint también informa en profundidad, aunque no tanto de los actores como del uso. Las “transparencias” ppt son hoy en día el medio más universal de presentar información en aulas, seminarios o congresos, allí donde exista una audiencia a la que se pretenda comunicar de forma rápida y rigurosa un conocimiento ya asentado o en vías de desarrollo.

Tabla IV
Universidades con más ficheros ppt (enero 2006)

	<i>DOC. MS POWERPOINT (ppt)</i>
UNIVERSIDAD DE SAO PAULO	19.600
UNIVERSIDAD ESTATAL DE CAMPINAS	10.800
UNIVERSIDAD DE LAS AMERICAS PUEBLA	1.850
UNIVERSIDAD DE VALE DO ITAJAI	1.320
PONTIFICIA UNIVERSIDAD CATOLICA DE RIO GRANDE DO SUL	928
PONTIFICIA UNIVERSIDAD CATOLICA DE PARANA	923
UNIVERSIDAD FEDERAL DE PERNAMBUCO	900
TECNOLOGICO DE MONTERREY	793
UNIVERSITY OF THE WEST INDIES	773
UNIVERSIDAD FEDERAL DE SANTA CATARINA	755

Sin embargo, un fichero ppt es fácilmente editable y la información se puede cortar y reutilizar de forma muy sencilla. Ello genera inquietud entre los autores por el peligro de plagio y hace que no sea habitual su publicación en abierto. Esto explicaría que las instituciones líderes según la tabla IV incluyan numerosas «novedades» y universidades claramente volcadas en la enseñanza no presencial. Cabría indicar que un número no precisado de presentaciones Powerpoint se distribuye en realidad como ficheros Acrobat (pdf) lo que refuerza la explicación sobre la «inseguridad» del formato como causa para su bajo nivel de publicación.

5. Conclusiones

Un grupo reducido de universidades latinoamericanas compite en igualdad con instituciones del resto del mundo en lo que respecta a su compromiso con la publicación y diseminación de conocimiento a través de la Web. Se trata fundamentalmente de grandes instituciones nacionales de carácter público, aunque también hay una buena representación de universidades católicas. Destacan por número (entre las mejor clasificadas) las universidades brasileñas y mejicanas (Sao Paulo, UNAM), aunque también otros parámetros las sitúan en las primeras posiciones.

La comunicación informal a través de formatos ricos parece asumida en todo este grupo de universidades de élite, posiblemente por iniciativas individuales de auto-archivo de documentos o mediante la creación de repositorios institucionales o temáticos de artículos. Nuevamente son las universidades brasileñas y mejicanas las

que presentan un mayor número de documentos en formatos ricos tales como pdf o doc, lo que indica que efectivamente publican de manera abierta e intencionadamente parte de sus contenidos de alto contenido científico. Los ficheros ps son indicativos de una actividad específicamente científica o tecnológica, al ser el formato preferido de matemáticos, físicos o ingenieros: los valores encontrados en universidades tecnológicas o con importantes escuelas de ingeniería recomiendan que la difícil evaluación bibliométrica de estas disciplinas se complemente con indicadores Web.

Aunque los ficheros ppt tienen una doble utilidad, tanto para fines docentes como para presentaciones en actos científicos, su número es muy bajo en instituciones líderes, debido posiblemente a reticencias sobre la propiedad intelectual, lo que debiera ser objeto de preocupación por parte de las autoridades académicas.

Es posible que las políticas e iniciativas que promueven la publicación en la Web no penetren todavía a todos los niveles académicos, pero ya son responsables de contribuir significativamente a la mejora de posición en Internet, atrayendo visitas y enlaces externos.

Bibliografía

- AGUILLO, I. F. (2002). Measuring Informal Scientific Publication in the Web. EASST 2002 Conference. International Conference of the European Association for the Study of Science and Technology. Universidad de York, Reino Unido, 31 Julio-3 de Agosto de 2002.
- 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=15agui12.htm>
- AGUILLO, I. F.; GRANADINO, B.; ORTEGA, J. L.; PRIETO, J. A. (2005). What the Internet says about science. *Scientist*, 19 (14): 10
- AGUILLO, I. F.; GRANADINO, B.; LLAMAS, G. (2005). Posicionamiento en el Web del sector académico iberoamericano. *Interciencia*, 30 (12): 1-5.
- AGUILLO, I. F.; ORTEGA, J. L.; PRIETO, J. A.; GRANADINO, B. (2005). Medida de la actividad y comunicación científica mediante indicadores cibernéticos. I Jornadas de Indicadores para la Evaluación de la Ciencia y la Tecnología. Madrid, 15 de abril de 2005. <<http://www.cindoc.csic.es/info/fesabid-prog.html>>
- AGUILLO, I. F.; GRANADINO, B.; ORTEGA, J. L.; PRIETO, J. A. (2006). Scientific research activity and communication measured with cybermetric indicators. *Journal of the American Society for Information Science and Technology*, 57 (10): 1296-1302.
- BAR-ILAN, J. (2005). Expectations versus reality – Search engine features needed for Web research at mid 2005. *Cybermetrics*, 9 (1), paper 2.
<<http://www.cindoc.csic.es/cybermetrics/articles/v9i1p2.html>>
- BJORNEBORN, L.; INGWERSEN, P. (2004). Towards a Basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55 (14): 1216-1227.

- HARNAD, S.; BRODY, T. (2004). Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals, *D-Lib Magazine*, 10 (6) June 2004. <<http://www.dlib.org/dlib/june04/harnad/06harnad.html>>
- SWAN, A. (2005). Open access self-archiving: An introduction. Key Perspectives Ltd, Truro (Reino Unido). <<http://eprints.ecs.soton.ac.uk/11006/01/jiscsum.pdf>>
- THELWALL, M. (2003). Web use and peer interconnectivity metrics for academic Web sites. *Journal of Information Sciences*, 29(1): 11-20.
- 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.

ANEXO

Tabla I
Tamaño y visibilidad

<i>País</i>	<i>N.º universidades</i>	<i>Tamaño</i>	<i>Visibilidad</i>
BRASIL	133	51.476	27.971
MÉXICO	76	44.530	25.166
PUERTO RICO	18	34.559	19.849
CHILE	50	36.420	18.366
JAMAICA	3	13.229	15.841
ARGENTINA	60	18.152	14.042
COSTA RICA	12	20.824	11.192
GUATEMALA	4	21.287	10.569
VENEZUELA	18	21.413	9.955
PARAGUAY	1	8.460	7.320
COLOMBIA	47	37.164	7.211
PERÚ	24	15.409	7.167
TRINIDAD Y TOBAGO	1	16.030	6.761
PANAMÁ	3	8.365	6.711
GRANADA	1	3.110	4.657
BOLIVIA	8	7.565	4.071
CUBA	5	8.372	4.021
ECUADOR	11	8.144	3.890
NICARAGUA	3	8.425	3.789
DOMINICA	1	1.766	3.384
EL SALVADOR	4	8.413	3.309
URUGUAY	11	9.165	3.233
REPÚBLICA DOMINICANA	4	7.173	2.854
ISLAS CAIMÁN	1	229	2.565
HONDURAS	1	2.773	1.745

Tabla II
Ficheros ricos

<i>País</i>	<i>N.º univer- sidades</i>	<i>pdf</i>	<i>doc</i>	<i>ppt</i>	<i>ps</i>
BRASIL	133	11.347	3.199	335	344
MÉXICO	76	9.468	895	117	166
CHILE	50	5.595	955	110	272
ARGENTINA	60	5.430	1.203	51	70
VENEZUELA	18	4.853	295	73	46
COSTA RICA	12	3.465	1.000	48	89
URUGUAY	11	3.150	167	68	55
COLOMBIA	47	2.787	369	74	25
PERÚ	24	2.877	188	49	0
PUERTO RICO	18	2.618	219	109	17
TRINIDAD Y TOBAGO	1	2.010	393	134	0
NICARAGUA	3	1.193	24	14	0
JAMAICA	3	330	338	318	0
GUATEMALA	4	669	204	31	3
CUBA	5	315	493	40	11
PARAGUAY	1	465	155	11	0
REPÚBLICA DOMINICANA	4	288	230	21	0
ECUADOR	11	337	132	23	0
BOLIVIA	8	213	231	13	9
PANAMA	3	312	31	8	0
EL SALVADOR	4	185	102	50	0
GRANADA	1	107	14	0	0
HONDURAS	1	55	0	9	0
DOMINICA	1	56	0	0	0
ISLAS CAIMÁN	1	23	9	0	0