



Jornadas de intercambio y reflexión acerca de la investigación en Bibliotecología

29 y 30 de Octubre de 2015

Mesa 4: Métricas de información científico/técnica

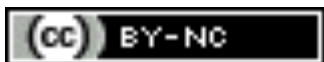
Poniendo en contexto los datos de descargas de documentos en repositorios digitales. Una aproximación a través del uso de percentiles

Cristian Merlino-Santesteban

Facultad de Cs. Económicas y Sociales
Universidad Nacional de Mar del Plata, Argentina
csantest@mdp.edu.ar

Esta obra se distribuye bajo licencia Creative Commons (CC) 3.0

http://creativecommons.org/licenses/by-nc/3.0/es/deed.es_AR



Repositorios digitales

Plataformas tecnológicas (abiertas) utilizadas en el proceso de democratización y socialización de los resultados de investigación para difundir y visibilizar la producción generada por universidades, centros de ciencia y tecnología y otras organizaciones de I+D.

Son valorados muy positivamente por su contribución a maximizar el alcance de la investigación científica y académica en acceso abierto.

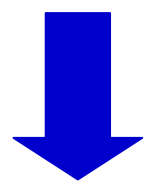


Construcción Valorativa



Juicios subjetivos

Juicios objetivos



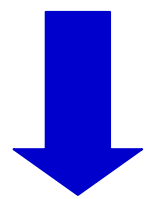
Evidencia empírica



Registros de transacciones de archivos (*webserver log files*)



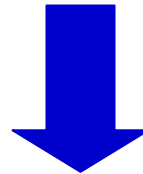
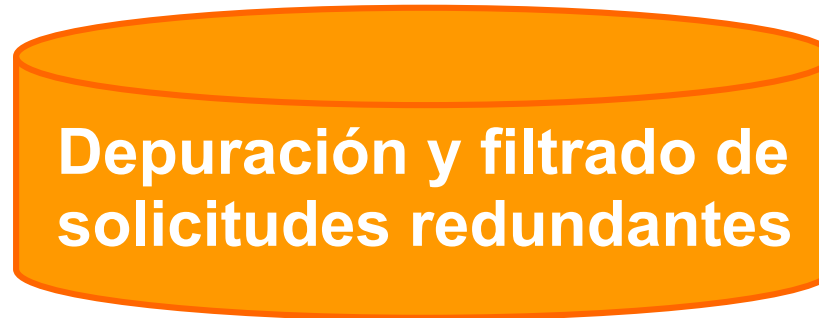
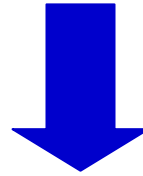
No es la única evidencia existente



Datos de acceso a objetos digitales dispuestos en abierto



Datos de acceso a objetos
digitales dispuestos en abierto



Cómputo del número de veces que un documento cualesquiera
fue descargado durante un período determinado



Grado de atención, difusión, visibilidad
o uso alcanzado



Presentación del uso de documentos

Por lo general, los datos de acceso a los objetos digitales son presentados en forma de rankings basados en conteos.

Los rankings de uso listan en orden descendente las publicaciones más descargadas o agregados significativos de éstas en función del número absoluto de descargas recibidas durante un espacio temporal determinado.

... permiten visualizar rápidamente los documentos **top**



LogEc

RePEc. Research Papers in Economics

Top 25 Journal Articles by File Downloads 2015-09

Rank	Journal Article	Click on a column heading to sort by a different category							
		File Downloads				Abstract Views			
		2015 09	3 months	12 months	Total	2015 09	3 months	12 months	Total
1	Introduction to Linear Regression Analysis, 5th edition by MONTGOMERY, DOUGLAS C., PECK, ELIZABETH A., and VINING, G. GEOFFREY <i>Hee-Seok Oh</i>	176	304	488	551	303	537	826	944
2	The corporate governance lessons from the financial crisis <i>Grant Kirkpatrick</i>	124	140	388	572	188	247	774	1,600
3	Human Resource Management, 11th Edition International Student Version. By David A. DeCenzo, Stephen P. Robbins and Susan L. Verhulst, Wiley, 2013; 448 Pages. Price £54.99 / €66.00, ISBN 978-1-1183-7971-4 <i>Shu-Kun Lin</i>	85	165	386	465	195	395	932	1,226
4	Henderson (James H.), Quandt (Richard E.) - Microeconomic theory: a mathematical approach <i>Capet Marcel</i>	77	189	783	2,060	189	438	1,705	4,405
5	Consumer Culture Theory (CCT): Twenty Years of Research <i>Eric J. Arnould and Craig J. Thompson</i>	75	121	449	2,977	193	357	1,367	7,888
6	Theory of the firm: Managerial behavior, agency costs and ownership structure <i>Michael C. Jensen and William H. Meckling</i>	66	162	912	7,522	184	509	3,031	19,560
7	The Role of Information Technology on the Banking Industry <i>Gabriela Cornelia Piciu and Georgiana Chitiga</i>	64	201	807	1,739	505	1,534	5,563	10,945
8	The Internationalization Process of the Firm—A Model of Knowledge Development and Increasing Foreign Market Commitments <i>Jan Johanson and Jan-Erik Vahlne</i>	62	100	468	2,574	150	318	1,555	7,444
9	Crime and Punishment: An Economic Approach <i>Gary S. Becker</i>	61	160	728	2,271	202	516	2,191	6,889
10	Tourism an Important Sector of Economy Development <i>Xhiliola Agaraj and Merita Murati</i>	59	130	384	770	1,535	3,395	10,203	17,185
11	Money creation in the modern economy <i>Michael McLeay, Amar Radia and Ryland Thomas</i>	58	62	86	92	241	273	360	392
12	Finance and Growth: Schumpeter Might Be Right <i>Robert G. King and Ross Levine</i>	57	167	766	10,449	141	365	1,818	27,935
12	Impact of Foreign Direct Investment on Indian Economy <i>Rangappa E.</i>	57	120	555	1,032	679	1,480	5,679	9,971
14	Econometric policy evaluation: A critique <i>Robert E. Lucas</i>	55	115	534	7,014	135	307	1,394	14,363

LogEc

RePEc. Research Papers in Economics

[Top Working Papers](#)

[Top Journal Articles](#)

[Top Books](#)

[Top Chapters](#)

[Top Software Items](#)


[Top Authors](#)

[Top Working Paper series and Journals](#)

Top 25 Working Papers by File Downloads 2015-09

Rank	Working Paper	Click on a column heading to sort by a different category							
		File Downloads				Abstract Views			
		2015 09	3 months	12 months	Total	2015 09	3 months	12 months	Total
1	Portfolio Rebalancing in Japan: Constraints and Implications for Quantitative Easing <i>Serkan Arslanalp and Dennis P. J. Botman</i>	353	353	353	353	987	987	987	987
2	The role of education in economic development: a theoretical perspective <i>Ilhan Ozturk</i>	247	769	3,031	6,836	2,445	7,174	25,880	56,711
3	Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data <i>Emmanuel Saez and Gabriel Zucman</i>	199	199	199	199	0	0	0	0
4	Cross-Licensing and Competition <i>Doh-Shin Jeon and Yassine Lefouill</i>	160	295	311	311	1	4	7	7
4	Whither Business History?: Memory, Message and Meaning <i>David Merrett</i>	160	160	160	160	4	5	5	5
6	Does EU Regulation Hinder or Stimulate Innovation? <i>Jacques Pelkmans and Andrea Renda</i>	148	393	434	434	8	10	18	18
7	Innovation and Legal Enforcement for Competition Policy: Theory and international evidence from overseas subsidiaries of the Japanese auto-parts suppliers <i>Yosuke Takeda and Ichihiro Uchida</i>	146	382	420	420	7	19	30	30
8	Denying Leniency to Cartel Instigators: Costs and Benefits <i>Zhiqi Chen, Subhadip Ghosh and Thomas W. Ross</i>	138	376	421	421	2	2	12	12
9	The Effects of Foreign Direct Investment on the Host Country Economic Growth - Theory and Empirical Evidence <i>Rui Moura and Rosa Portela Forte</i>	137	386	1,741	4,531	745	2,066	9,713	25,321
10	Harmonising Hayek and Posner: revisiting Posner, Hayek & the economic analysis of Law <i>Marianne Olo</i>	134	258	293	293	1	6	10	10
11	On the Theoretical Efficacy of Quantitative Easing at the Zero Lower Bound <i>Paola Boel and Christopher Waller</i>	126	126	126	126	1	1	1	1
12	Roots of Autocracy <i>Oded Galor and Marc Patrick Brag Klemp</i>	119	119	119	119	61	61	61	61
13	F. A. Hayek and the economic calculus <i>Bruce Caldwell</i>	115	140	140	140	1	1	1	1
14	MATLAB for Economics and Econometrics A Beginners Guide <i>John C. Frain</i>	112	239	437	437	191	403	669	669
15	The Greek Debt Crisis: Suggested Solutions and Reforms <i>Leo Michellis</i>	111	563	859	1,206	346	2,150	3,277	4,226

Social Science Research Network

SSRN Top 10,000 Papers 

Updated Weekly - Last Updated on: 10/18/2015

Ranked by: Total New Downloads

Search Papers

Rank	Paper	Last 12 Months		All Time				Authors							
		Total New Downloads		Total # of Downloads		Total # of Citations		# of Authors	Total Downloads per author		New Downloads per author		Total Citations per author		
1	How Fox News Changed American Media and Political Dynamics Bruce Bartlett Independent Date posted to database: May 11, 2015 Last Revised: June 4, 2015	31,411	(1)	31,411	(32)	0	(166746)	1	(246081)	31,411	(16)	31,411	(1)	0	(155292)
2	A Brief Introduction to the Basics of Game Theory Matthew O. Jackson Stanford University - Department of Economics Date posted to database: December 5, 2011 Last Revised: September 26, 2015	29,078	(2)	104,761	(4)	0	(166746)	1	(246081)	104,761	(3)	29,078	(2)	0	(155292)
3	A Quantitative Approach to Tactical Asset Allocation Mebane T. Faber Cambria Investment Management Date posted to database: February 11, 2007 Last Revised: March 3, 2014	29,073	(3)	162,643	(1)	15	(23998)	1	(246081)	162,643	(1)	29,073	(3)	15	(13531)
4	'A Diamond Is Forever' and Other Fairy Tales: The Relationship between Wedding Expenses and Marriage Duration Andrew M. Francis and Hugo M. Mialon Emory University - Department of Economics and Emory University - Department of Economics Date posted to database: September 27, 2014 Last Revised: September 27, 2014	17,952	(4)	34,462	(23)	0	(166746)	2	(99490)	17,231	(41)	8,976	(5)	0	(155292)
5	A Five-Factor Asset Pricing Model Eugene F. Fama and Kenneth R. French University of Chicago - Finance and Dartmouth College - Tuck School of Business Date posted to database: June 30, 2013 Last Revised: September 23, 2014	12,641	(5)	22,070	(58)	2	(90277)	2	(99490)	11,035	(107)	6,321	(14)	1	(93450)
6	Understanding the Modern Monetary System Cullen O. Roche Orcam Financial Group, LLC Date posted to database: August 6, 2011 Last Revised: July 8, 2015	12,337	(6)	55,524	(7)	0	(166746)	1	(246081)	55,524	(5)	12,337	(4)	0	(155292)

UCL Discovery download statistics

▼ Top 20 downloads last month: April 2015

Rank	Paper	Downloads
1	Pheng, L.S.; (1990) Marketing theories and concepts for the international construction industry: a study of their applicability at the global, national and corporate perspectives. Doctoral thesis, University of London.	849
2	Brennan, F.P.; (1992) Fatigue and fracture mechanics analysis of threaded connections. Doctoral thesis, University of London.	761
3	Latif, A.; (1999) A study of the design of fluidized bed reactors for biomass gasification. Doctoral thesis, University of London.	739
4	Liu, Q.; (2010) Efficiency analysis of container ports and terminals. Doctoral thesis, UCL (University College London).	707
5	Kassem, A.H.; (2006) The legal aspects of seaworthiness: current law and development. Doctoral thesis, Swansea University.	582
6	Marmot, M; Atinmo, T; Byers, T; Chen, J; Hirohata, T; Jackson, A; ... Zeisel, S; + view all (2007) Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. (WCRF/AICR Expert Report). World Cancer Research Fund / American Institute for Cancer Research: Washington DC, US.	542
7	Issa, N.; (2011) Towards more efficient screening of pharmaceutical cocrystals. Doctoral thesis, UCL (University College London).	496
8	Keval, H.U.; (2009) Effective design, configuration, and use of digital CCTV. Doctoral thesis, UCL (University College London).	494
9	Fletcher, T.S.B.; (2012) Machine learning for financial market prediction. Doctoral thesis, UCL (University College London).	467
10	Sharp, H.; Finkelstein, A.; Galal, G.; (1999) Stakeholder identification in the requirements engineering process. In: Proceedings of 10th International Workshop on Database & Expert Systems Applications (DEXA). (pp. pp. 387-391). IEEE Computer Society Press	461
11	Rozum, I.; Tennyson, J.; (2004) Elastic differential cross sections for the CF_x (x = 1, 2, 3) radicals. Journal of Physics B: Atomic, Molecular and Optical Physics , 37 (5) pp. 957-966.	451



DIGITAL ACCESS TO
SCHOLARSHIP AT HARVARD

Works Downloaded

Downloads Per Work

(School/Department/Collection)

Harvard

Site usage reports

- Journal requests
- Articles requests by language
- Issue requests
- Article requests

* Count started in May , 2007.

Initial Date: Final Date:

Display issue visited times or more.

nr. of requests	article reference
22790	MATTOS, Fernando Augusto Mansor de and CHAGAS, Gleison José do Nascimento. Challenges for digital inclusion in Brazil. <i>Perspect. ciênc. inf.</i> , 2008, vol.13, no.1, p.67-94. ISSN 1413-9936. · abstract in portuguese english · text in portuguese
19660	CUNHA, Murilo Bastos da. From conventional to digital libraries: differences and convergences. <i>Perspect. ciênc. inf.</i> , 2008, vol.13, no.1, p.2-17. ISSN 1413-9936. · abstract in portuguese english · text in portuguese
17463	GUIMARÃES, Lúcia Filomena de Almeida and FALSARELLA, Orandi Mina. An analisys of Just-in-Time methodology and Kanban production system under the information science point of view. <i>Perspect. ciênc. inf.</i> , 2008, vol.13, no.2, p.130-147. ISSN 1413-9936. · abstract in portuguese english · text in portuguese
14870	RODRIGUES, Ana Márcia Lutterbach. Archival theory and records management. <i>Perspect. ciênc. inf.</i> , 2006, vol.11, no.1, p.102-117. ISSN 1413-9936. · abstract in portuguese english · text in portuguese
14472	CASTRO, César Augusto and SOUSA, Maria Conceição Pereira de. Project's pedagogy in school library: proposal of a model for the process of school research. <i>Perspect. ciênc. inf.</i> , 2008, vol.13, no.1, p.134-151. ISSN 1413-9936. · abstract in portuguese english · text in portuguese
14267	CAPURRO, Rafael and HJORLAND, Birger. The concept of information as we use in everyday. <i>Perspect. ciênc. inf.</i> , 2007, vol.12, no.1, p.148-207. ISSN 1413-9936. · abstract in portuguese english · text in portuguese

(Simples) Conteos

Los conteos brutos sin normalizar provistos por los ordenamientos permiten discernir fácilmente entre los documentos que captaron mayor atención y aquellos que no, pero poco informan sobre el contexto de análisis y menos aún sobre la posición relativa de cada publicación en el conjunto de datos estudiados.

¿Y si relativizamos los recuentos?

Los R basados en promedios de descargas se ven afectados por el comportamiento asimétrico de las distribuciones de descargas.

Comportamiento de las distribuciones de datos de descargas

Las distribuciones de datos de descargas se caracterizan por:

ser muy asimétricas y

seguir un comportamiento similar a las distribuciones de datos de citaciones a documentos, aunque no tan acusado.



Asimetría

Se dice que una distribución es asimétrica cuando hay presencia de valores atípicos (valores muy altos o muy bajos), es decir, hay observaciones muy alejadas de las demás las que ocasionan una mayor dispersión en los datos de la distribución.



Limitaciones de los indicadores de uso basados en simples recuentos

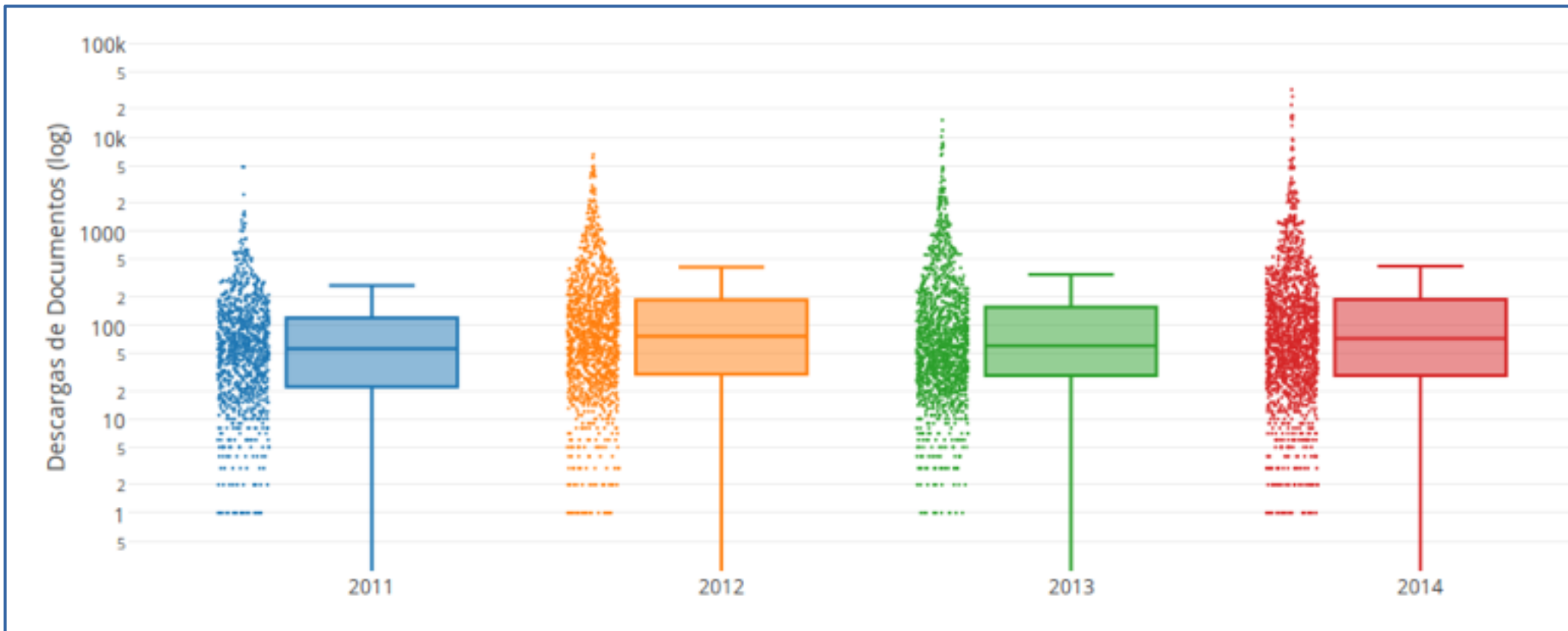
Padecen de las mismas debilidades que se les atribuyen a los indicadores bibliométricos básicos basados en conteos de citas.

Los recuentos absolutos son poco informativos por sí mismo si no se consideran las especificidades propias de cada campo disciplinar y se establece un estándar de referencia con el que confrontar tales conteos.

Los valores medios de descargas se ven seriamente afectados por algunos valores extremos, alejándolos del patrón general.



Distribución de descargas de documentos de un repositorio digital. Años 2011, 2012, 2013 y 2014



Estadística descriptiva. Descarga de documentos

Año 2011	
Media	116,26
Moda	1
Mediana	56
Primer cuartil	22
Tercer cuartil	119
Desviación estándar	269,35
Mínimo	0
Máximo	4.839
Suma	129.402
Conteo	1.113

Año 2012	
Media	227,03
Moda	0
Mediana	76
Primer cuartil	30
Tercer cuartil	185
Desviación estándar	550,30
Mínimo	0
Máximo	6.529
Suma	301.499
Conteo	1.328



Estadística descriptiva. Descarga de documentos

Año 2013	
Media	257,65
Moda	20
Mediana	60
Primer cuartil	29
Tercer cuartil	155
Desviación estándar	871,21
Mínimo	0
Máximo	15.179
Suma	394.979
Conteo	1.533

Año 2014	
Media	357,81
Moda	0
Mediana	72
Primer cuartil	29
Tercer cuartil	187
Desviación estándar	1.616,42
Mínimo	0
Máximo	32.012
Suma	622.945
Conteo	1.741



Cómo podemos mejorar la comparación

Necesidad de medidas (de posición/localización) que caractericen de mejor manera cada publicación en relación a su población de referencia ==> Los percentiles son una de esas medidas.

A través de ellos podemos indicar la posición relativa que ocupa un determinado documento en relación a todas las publicaciones que conforman el grupo de referencia o control.



Qué son los percentiles I

Los percentiles (o centiles) son los cuantiles que dividen la serie de datos en 100 partes iguales, cada una de ellas con el 1% de los datos. El percentil 100 es el mayor valor del set y el percentil 0 es el menor valor. De esta forma podemos decir genéricamente que un determinado valor se sitúa en el percentil 80 (P_{80}) cuando deja por debajo de sí el 80% de las observaciones de la población y simultáneamente es superado por el 20% restante.



Qué son los percentiles II

Cuando se expresa que un documento se sitúa en el percentil 90 (P_{90}) con respecto a su nivel de descarga, se está diciendo que supera al 90% de las publicaciones del conjunto de referencia.

El percentil 50 (P_{50}) es la mediana, por lo tanto es el umbral de descargas que identifica la repercusión promedio.



Uso de percentiles. Ventajas

No son tan influenciados por los valores anómalos y se pueden aplicar a distribuciones no normales.

Estas medidas de localización no son tan sensibles como la media aritmética ni tan insensibles como la mediana ante la presencia de colas.



Advertencia

Para que las contextualizaciones sean verdaderamente significativas los documentos que conforman el grupo de referencia deben ser “similares” a la publicación a comparar.

Se ha advertido que características como la tipología documental y la temática afectan el patrón de descarga de los documentos, introduciendo una alta variabilidad.



Demostración práctica

La Tabla 1 presenta una distribución ficticia de bajadas de documentos para un espacio temporal X. Las publicaciones se encuentran ordenadas ascendentemente por el número de descargas. A los documentos con el mismo número de bajadas se les asignó el rango promedio. Cada publicación fue asignada a un percentil según su rango usando la formula: $(100 * (i-0,5) / n)$. Los documentos con cero descargas fueron asignados al percentil 0.

n	i	D	P
50	50	10.508	99,00
49	49	7.233	97,00
48	48	4.567	95,00
47	47	4.323	93,00
46	46	3.122	91,00
45	45	804	89,00
44	43,5	702	86,00
43	43,5	702	86,00
42	42	601	83,00
41	41	589	81,00
40	40	521	79,00
39	38	405	75,00
38	38	405	75,00
37	38	405	75,00
36	36	322	71,00
35	35	313	69,00
34	32	278	63,00
33	32	278	63,00
32	32	278	63,00
31	32	278	63,00
30	32	278	63,00
29	29	134	57,00
28	28	132	55,00

n	i	D	P
27	27	121	53,00
26	23	109	45,00
25	23	109	45,00
24	23	109	45,00
23	23	109	45,00
22	23	109	45,00
21	23	109	45,00
20	23	109	45,00
19	19	86	37,00
18	18	76	35,00
17	17	45	33,00
16	16	43	31,00
15	15	37	29,00
14	14	36	27,00
13	12	35	23,00
12	12	35	23,00
11	12	35	23,00
10	10	24	19,00
9	9	15	17,00
8	8	4	15,00
7	7	0	0
6	6	0	0
5	5	0	0

A modo de cierre

Necesidad de introducir métricas estadísticamente más fiables que permitan, dentro de un marco comparativo común, normalizar los datos de uso a fin de obtener mejores comparaciones.

Los indicadores basados en percentiles son una solución posible, ya que nos permiten comparar de mejor manera el desempeño de un documento con respecto a un set de referencia.

