

# Cybermetric indicators

## A methodological approach 2009

**Isidro F. Aguillo**

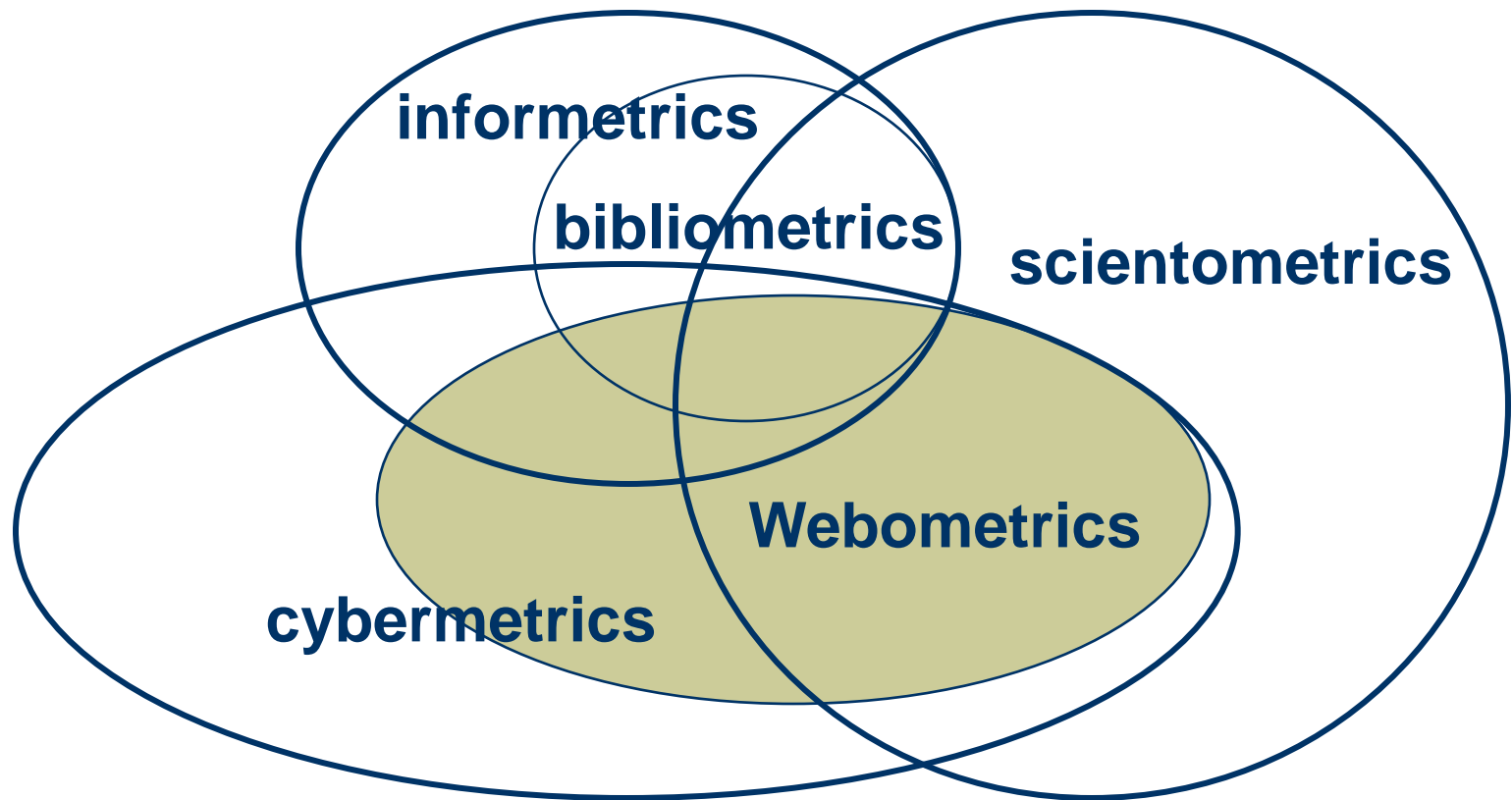
Cybermetrics Lab  
*isidro@cindoc.csic.es*

**2<sup>nd</sup> International Workshop on University Web Rankings**  
*CCHS-CSIC, Madrid (Spain). April 21<sup>st</sup> 2009*

# Introduction

- ◆ Cybermetrics is devoted to the quantitative description of the contents and communication activities that occurs in the cyberspace
  - For practical reasons Cyberscientometrics is the most developed sub-field and it is simply called Cybermetrics
  - As most of the academic related material are published on the Web usually the preferred term is **Webometrics**

# Quantitative disciplines

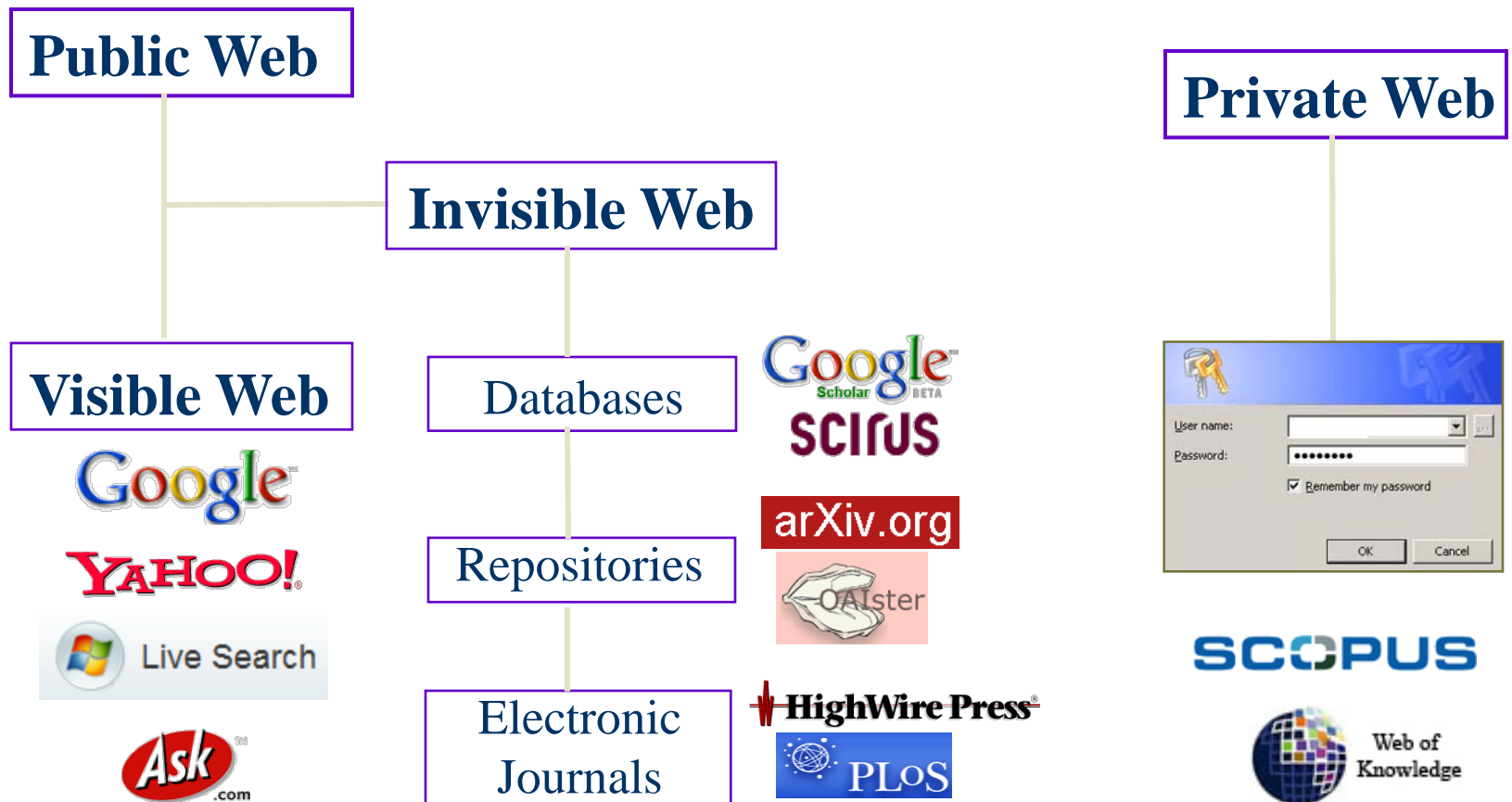


Adapted from Björneborn

# Contents

<b>CYBERSPACE (Contents in electronic format)</b>				
<b>I N T E R N E T</b>	<b>CONTENTS</b>	<b>OPEN (PUBLIC) INTERNET</b>	<b>EMAIL, FORUMS, USENET NEWS</b>	
			<b>WEBSPACE</b>	<b>VISIBLE WEB</b>
				<b>INVISIBLE WEB (DEEP WEB)</b>
			<b>INFRANET</b>	
			<b>DATA ABOUT INTERNET USAGE</b>	
	<b>INTRANET</b>			
	<b>PHYSICAL INTERNET DATA</b>	<b>TOPOLOGY, TRAFFIC, DEMOGRAPHY, GEOGRAPHY</b>		
<b>OUTSIDE INTERNET</b>				

# Web



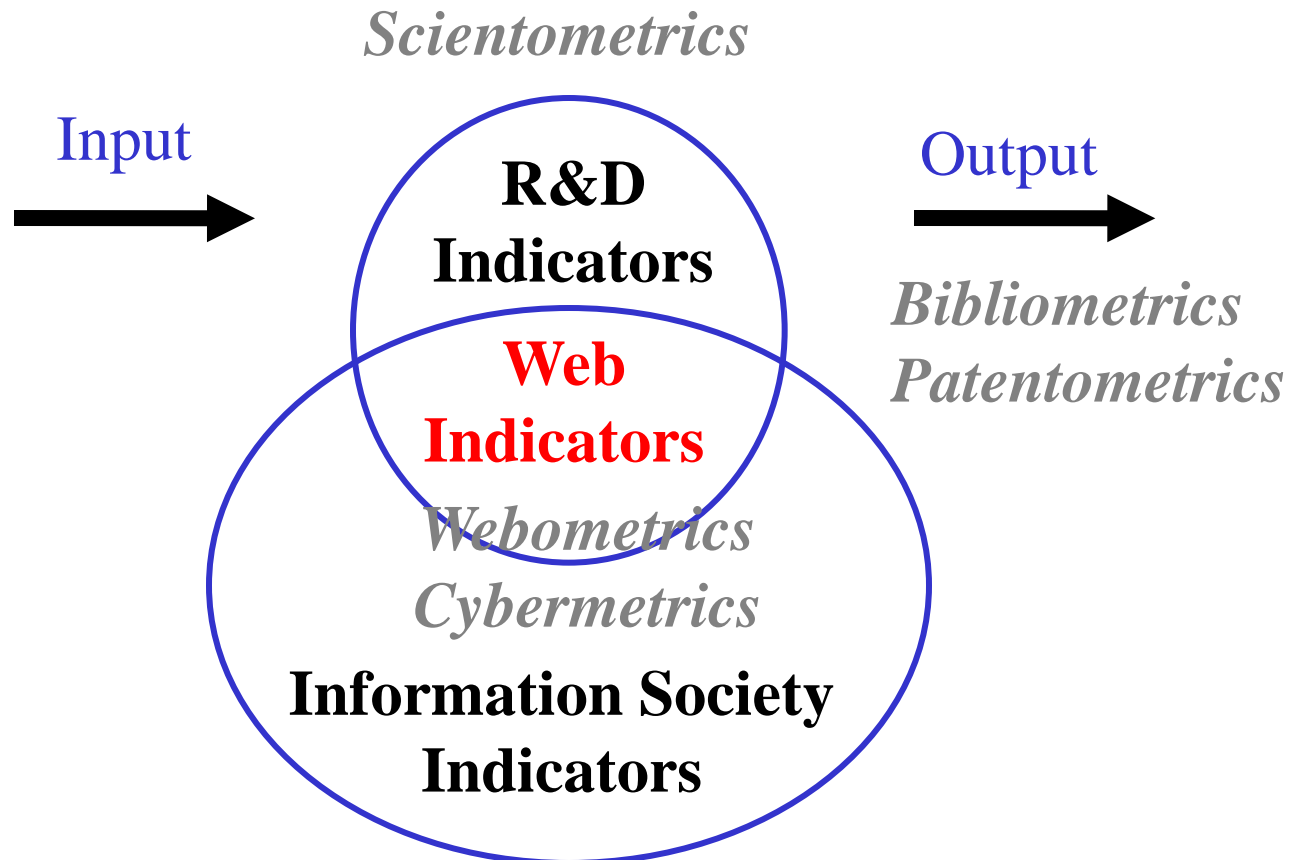
# Advantages

- ◆ The presence on the Web of research groups, professors or postgraduate students reflects more activities and results than the traditional formal publication in refereed journals
  - unpublished material, general public contributions, drafts for future papers or book chapters, slides used in conference or seminar presentations, support material for courses or even raw data
- ◆ The Web reaches a wider audience than the paper based publications like journals or books.
  - The information published on the Web can be recovered by any Internet user worldwide
- ◆ The interlinked nature of the Web offers the possibility to discover hidden relationships among different websites
  - Identifying academic communities but showing also economic, industrial, social or cultural relationships

# I

Describing academic and  
research activities using Web  
search engines as new  
intermediaries

# Web indicators



# Counting websites

- ◆ Direct method: crawler/robot
  - Software libraries already available (free/low cost)
  - Technical knowledge needed for customising robots
  - Fine detail customisation needs lots of tests
  - Not as good as robots of commercial search engines
  - Huge computer resources (and time) needed
- ◆ Indirect method: search engines
  - Universal access
  - Strong performance
  - Commercial secrets
  - Irregular behaviour

# Xenu LinkSleuth

Address	Status	Type	Size	Title
http://www.cindoc.csic.es/	ok	text/html	43785	CINDOC-Centro de Informacion y Documentacion Cientifica
http://www.csic.es/	ok	text/html	84	
http://www.cindoc.csic.es/images/logo_csic.jpg	ok	image/jpeg	3383	
http://www.cindoc.csic.es/images/logo-67.gif	ok	image/gif	4395	
http://www.cindoc.csic.es/images/cabecera.jpg	ok	image/jpeg	72917	
http://www.cindoc.csic.es/info/accesos.html	ok	text/html	26514	CINDOC-Contacto y Accesos
http://www.cindoc.csic.es/images/menu/contacto1.jpg	ok	image/jpeg	3718	
http://www.cindoc.csic.es/tarifas.html	ok	text/html	41225	CINDOC--Tarifas
http://www.cindoc.csic.es/images/menu/tarifas1.jpg	ok	image/jpeg	3230	
http://www.cindoc.csic.es/info/mapa.html	ok	text/html	30116	
http://www.cindoc.csic.es/images/menu/mapa1.jpg	ok	image/jpeg	3319	
http://www.cindoc.csic.es/images/ico_casa.gif	ok	image/gif	77	
http://www.cindoc.csic.es/eng/principal.html	ok	text/html	45202	
http://www.cindoc.csic.es/images/gb.gif	ok	image/gif	218	
http://www.cindoc.csic.es/images/menu/elcindoc.jpg	ok	image/jpeg	7261	
http://www.cindoc.csic.es/info/infobjetivos.html	ok	text/html	32674	
http://www.cindoc.csic.es/images/menu/presentacion1.jpg	ok	image/jpeg	5544	
http://www.cindoc.csic.es/info/personal.html	ok	text/html	3583	
http://www.cindoc.csic.es/images/menu/personal1.jpg	ok	image/jpeg	4785	
http://www.cindoc.csic.es/info/memoria.pdf	ok	applicati...	24...	
http://www.cindoc.csic.es/images/menu/memoria1.jpg	ok	image/jpeg	4728	
http://www.cindoc.csic.es/images/menu/investigacion.jpg	ok	image/jpeg	7462	
http://www.cindoc.csic.es/investigacion/departamentos.html	ok	text/html	30882	
http://www.cindoc.csic.es/images/menu/departamentos1.jpg	ok	image/jpeg	5902	
http://www.cindoc.csic.es/investigacion/investigacion.html	ok	text/html	25198	
http://www.cindoc.csic.es/images/menu/lineas1.jpg	ok	image/jpeg	7388	
http://www.cindoc.csic.es/investigacion/grupos.html	ok	text/html	24838	
http://www.cindoc.csic.es/images/menu/grupos1.jpg	ok	image/jpeg	7502	
http://www.cindoc.csic.es/investigacion/proyectos.html	ok	text/html	40603	
http://www.cindoc.csic.es/images/menu/proyectos1.jpg	ok	image/jpeg	5042	
http://www.cindoc.csic.es/investigacion/informes.html	ok	text/html	24151	
http://www.cindoc.csic.es/images/menu/resultados1.jpg	ok	image/jpeg	5250	
http://www.cindoc.csic.es/images/menu/productos1.jpg	ok	image/jpeg	6820	
http://www.cindoc.csic.es/servicios/dbinfo.htm	ok	text/html	39892	
http://www.cindoc.csic.es/images/menu/database1.jpg	ok	image/jpeg	7243	
http://www.cindoc.csic.es/webpublic/publicac.htm	ok	text/html	29407	
http://www.cindoc.csic.es/images/menu/publicaciones1.jpg	ok	image/jpeg	5594	
http://www.cindoc.csic.es/isis/isis.htm	ok	text/html	8397	
http://www.cindoc.csic.es/images/menu/isis1.jpg	ok	image/jpeg	5017	

MIME type	count	Σ size	Σ size (KB)	% size
text/html	904 URLs	28015352 Bytes	(27358 KB)	51.13%
image/jpeg	156 URLs	1967472 Bytes	(1921 KB)	8.82%
image/gif	179 URLs	522343 Bytes	(510 KB)	10.12%
application/pdf	438 URLs	89064021 Bytes	(86976 KB)	24.77%
text/css	2 URLs	21706 Bytes	(21 KB)	0.11%
application/x-javascript	5 URLs	30189 Bytes	(29 KB)	0.28%
application/x-shockwave-flash	1 URLs	0 Bytes	(0 KB)	0.06%
application/msword	22 URLs	43891200 Bytes	(42862 KB)	1.24%
application/zip	2 URLs	21588994 Bytes	(21083 KB)	0.11%
image/png	50 URLs	0 Bytes	(0 KB)	2.83%
application/vnd.ms-powerpoint	9 URLs	6765056 Bytes	(6606 KB)	0.51%
<b>Total</b>	<b>1768 URLs</b>	<b>191866333 Bytes</b>	<b>(187369 KB)</b>	<b>100.00%</b>

# Units: institutional domain

DOMAIN	Top Level Domain	gTLD	<i>.com, .org, .net, .info, .name, .eu, .cat (.cc, .fm, .nu, .tv, .fm, ...)</i>
		cTLD	<i>.uk, .de, .es, .it, .be, .ca, .ru, .au, .za (.us, .mil, .gov, (.edu))</i>
		Subdomain	<i>.ac.uk, .ac.be, .ac.be, .edu.hk, .edu.tw</i>
	Institutional	Subdomain	<i>.harvard.edu, .cam.ac.uk, .ucm.es .insu.cnrs.fr, .ibp.cnr.it, .ipp.mpg.de</i>
SITE	Site		<i>www.cindoc.csic.es</i>
	Directory/Subdirectory		<i>www.cindoc.csic.es/cybermetrics</i>
	Document/Rich file		<i>.html, .asp, .php, .pdf, .doc, .ppt</i>

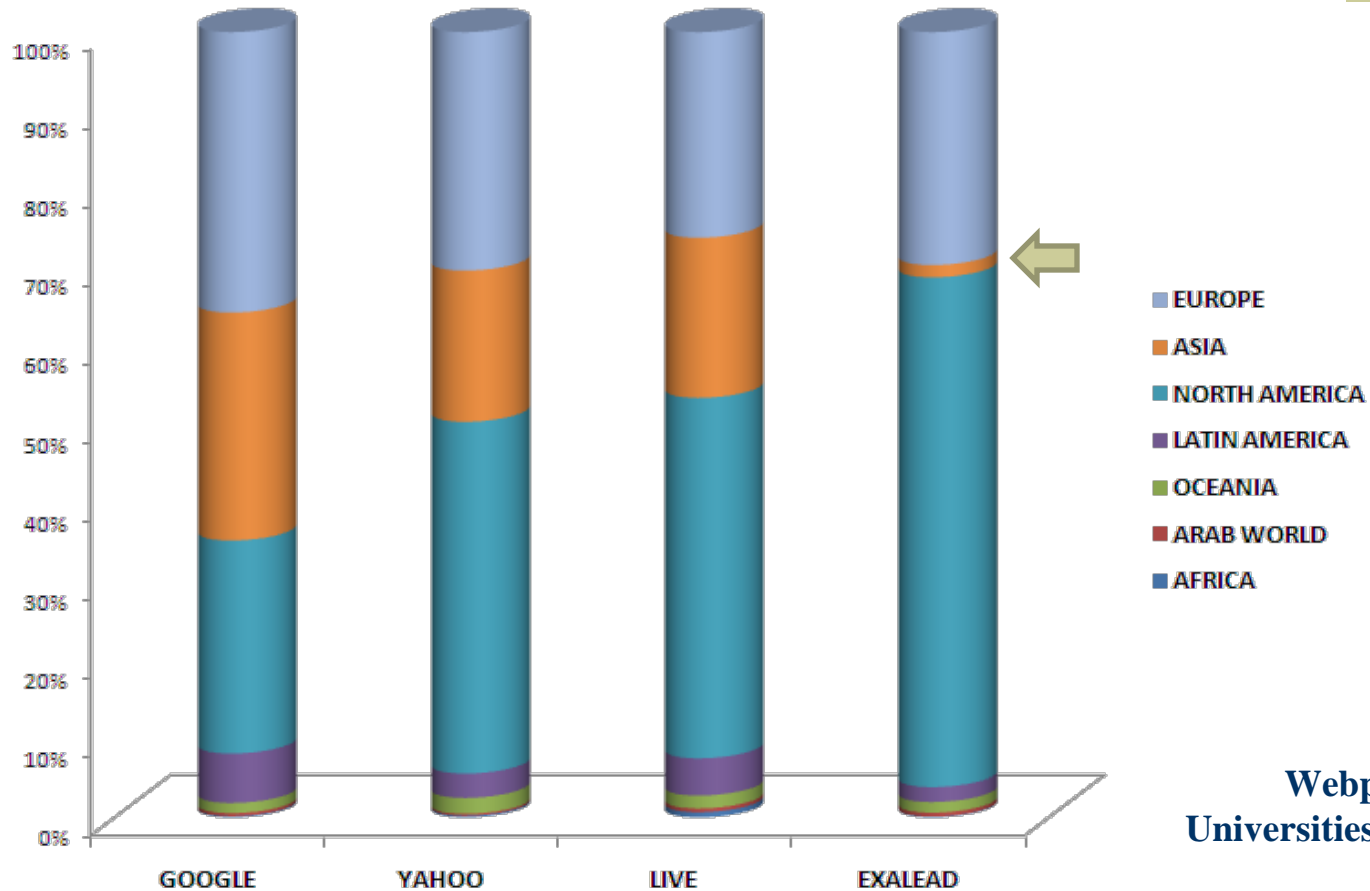
Name	Database			
	2003	2004-05	2006-07	2008
Altavista	ALTAVISTA	YAHOO		
Alltheweb	FAST			
Yahoo	GOOGLE			
Google		GOOGLE		
Alexa	»	MSN SEARCH	ALEXA	ALEXA
A9			LIVE (MSN)	
MSN Search		INKTOMI	ASK (TEOMA)	LIVE
Hotbot	INKTOMI	«		
Lycos	FAST	TEOMA	ASK (TEOMA)	ASK
Teoma	TEOMA			
Gigablast	GIGABLAST			
Exalead	»	EXALEAD		
Wisenut	WISENUT			«

# Methods

## Syntax of the main search engines

OPERATORS	GOOGLE	YAHOO	LIVE	EXALEAD	ASK	GIGABLAST
<b>Top Level Domain</b>	site:aa.xx	Site Explorer http://xx	site:aa.xx	site:aa.xx	site:aa.xx inurl:aa.xx	site:aa.xx
<b>Institutional Domain</b>	site:aa.xx	Site Explorer http://aa.xx	site:aa.xx	site:aa.xx	site:aa.xx inurl:aa.xx	site:aa.xx
<b>Directory</b>	site:aa.xx/bb	site:aa.xx inurl:bb	site:aa.xx/bb	site:aa.xx/bb	inurl:aa.xx/bb	site:aa.xx suburl:bb
<b>Term in URL</b>	inurl:bb	inurl:bb	inurl:bb	inurl:bb	inurl:bb	suburl:bb
<b>External</b>	NO	linkdomain:aa.xx	NO	link:aa.xx	NO	NO
<b>Country</b>	Advanced	Advanced	loc:XX	country:XX	Advanced	NO
<b>Language</b>	Advanced	Advanced	language:zz	language:zz	Advanced	Advanced
<b>File format</b>	filetype:yy	originurlextension:yy	filetype:yy	filetype:yy	NO	type:yy

# Geographical biases



Webpages from  
Universities (January, 2009)

[Advanced Search](#)  
[Preferences](#)

Web

Results 1 - 10 of about **276,000**

### [Careers](#)

Undergraduates. Undergraduate planning, help with CVs & interv  
[www.ic.ac.uk/careers/](http://www.ic.ac.uk/careers/) - 24k - C

### [SAS](#)

What's New: welcome to statist  
statistician, joins the SAS team  
[www.ic.ac.uk/stathelp/](http://www.ic.ac.uk/stathelp/) - 18k - C

### [History of Science](#)

centre for the history of science  
[www.hstm.ic.ac.uk/](http://www.hstm.ic.ac.uk/) - 22k - Cac

### [Chemical Engineering](#)

Welcome to the department of  
College London.  
[www.ce.ic.ac.uk/](http://www.ce.ic.ac.uk/) - 19k - Cache

[Advanced Search](#)  
[Preferences](#)

Web

Results 1 - 10 of about **129,000**

### [Careers](#)

Undergraduates. Undergraduates. Information, advice and guidance. Advice with career planning, help with CVs & interviews & applications. ...  
[www.imperial.ac.uk/careers/](http://www.imperial.ac.uk/careers/) - 24k - [Cached](#) - [Similar pages](#)

### [Electron Microscopy](#)

Electron Microscopy in the Department of Life Sciences.  
[www.imperial.ac.uk/electronmicroscopy/](http://www.imperial.ac.uk/electronmicroscopy/) - 16k - [Cached](#) - [Similar pages](#)

### [Summer Accommodation](#)

Summer Accommodation in Central London, South kensington, just step away from Science Museums, Natural History Museums and Hyde Park.  
[www.imperial.ac.uk/summeraccommodation/](http://www.imperial.ac.uk/summeraccommodation/) - 18k - [Cached](#) - [Similar pages](#)

### [The Centre for Development, Environment and Policy at SOAS](#)

The Centre for Development, Environment and Policy containing The Distance Learning Programme at Wye providing MSc and postgraduate qualifications in ...  
[www.imperial.ac.uk/distancelearning/](http://www.imperial.ac.uk/distancelearning/) - 13k - [Cached](#) - [Similar pages](#)

[Advanced Search](#)  
[Preferences](#)

Web

Results 1 - 10 of about **388,000**


# Google Dance

site:cnr.it Search Search:  NEW! Datacenters  64.233.\*  64.\*  72.\*  by C-Blocks Results: 10  III  III  III

Google Watch Tool CMC-Par Inc 2004  .99  .104  .107  .147  allins 64.233.\*.107 C-Blocks Window: Same GO

IP	Search Results	Time
.161	Resultados 1 - 10 de aproximadamente 691,000 de cnr.it	0.33 segundos
.163	Resultados 1 - 10 de aproximadamente 677,000 de cnr.it	0.26 segundos
.167	Resultados 1 - 10 de aproximadamente 677,000 de cnr.it	0.23 segundos
.169	Resultados 1 - 10 de aproximadamente 691,000 de cnr.it	0.28 segundos
.171	Resultados 1 - 10 de aproximadamente 691,000 de cnr.it	0.26 segundos
.179	Resultados 1 - 10 de aproximadamente 691,000 de cnr.it	0.26 segundos
.181	Resultados 1 - 10 de aproximadamente 707,000 de cnr.it	0.40 segundos

# Google Data Centers

IP/domain	Pages (thousands)	
72.14.207.99	1103979	3,64%
642.33.161.107	1127195	1,61%
66.102.1.107	1128276	1,51%
216.239.59.99	1156158	0,92%
642.33.179.104	1152480	0,60%
google.co.uk	1149023	0,30%
72.14.219.104	1147693	0,18%
209.85.133.104	1147691	0,18%
<b>google.com</b>	<b>1145627</b>	
66.249.89.107	1207208	5,38%
72.14.235.104	1227079	7,11%

Webpages from Universities  
according several  
Google Data Centers  
(January, 2008)

# Rounded results

**YAHOO! SEARCH** [Web](#) | [Images](#) | [Video](#) | [Audio](#) | [Directory](#) | [Local](#) | [News](#) | [Shopping](#) | [More »](#)  
site:cindoc.csic.es Search

**Search Results** 1 - 100 of about 22,700 from cindoc.csic.es

**YAHOO! SEARCH** [Web](#) | [Images](#) | [Video](#) | [Audio](#) | [Directory](#) | [Local](#) | [News](#) | [Shopping](#) | [More »](#)  
site:cindoc.csic.es Search

**Search Results** 101 - 200 of about 22,500 from cindoc.csic.es

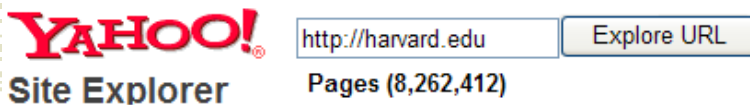
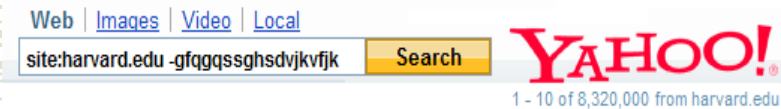
**YAHOO! SEARCH** [Web](#) | [Images](#) | [Video](#) | [Audio](#) | [Directory](#) | [Local](#) | [News](#) | [Shopping](#) | [More »](#)  
site:cindoc.csic.es Search

**Search Results** 401 - 500 of about 21,700 from cindoc.csic.es

**YAHOO! SEARCH** [Web](#) | [Images](#) | [Video](#) | [Audio](#) | [Directory](#) | [Local](#) | [News](#) | [Shopping](#) | [More »](#)  
site:cindoc.csic.es Search

**Search Results** 701 - 800 of about 20,100 from cindoc.csic.es

# Yahoo! Mirrors



Databases	Webpages Top 1000	
<b>YAHOO</b> ←	<b>38052600</b>	<b>4,01%</b>
YAHOO API	36556800	-0,08%
ALLTHEWEB	36539700	-0,12%
ALTAVISTA	36668200	0,23%
YAHOO SITE EXPLORER	35105196	-4,04%

Databases served by Yahoo: Webpages in the Top 1000 University domains (July, 2008)

# Yahoo SiteExplorer

**YAHOO!** SEARCH

## Site Explorer

[Site Explorer Home](#) - [Feedback](#) - [Blog](#) - [Badge](#) - [Preferences](#)

Results Inlinks **1 - 100** of about **547,043** - 0 sec.

[Pages \(527,015\)](#) | Inlinks **(548,000)**

Show Inlinks  to:

### 1. [Guide to Philosophy on the Internet \(Suber\)](#)

[www.earlham.edu/~peters/philinks.html](http://www.earlham.edu/~peters/philinks.html) [Web](#) | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#)

### 2. [Seismo-surfing the Internet - Exploring Earthquake Data](#)

[www.geophys.washington.edu/seismosurfing.html](http://www.geophys.washington.edu/seismosurfing.html)

**YAHOO!**

1 - 10 of 155,000 for linkdomain: csic. es- site: csic. es

### [Guide to Philosophy on the Internet](#)

A comprehensive collection of online philosophy resources. Well-organized, easy to navigate, regularly updated.

[www.earlham.edu/~peters/philinks.htm](http://www.earlham.edu/~peters/philinks.htm) - 380k - [Cached](#)

### [Surfing the Internet for Earthquake Data](#)

Index of earthquake-related sites.

[www.geophys.washington.edu/seismosurfing.html](http://www.geophys.washington.edu/seismosurfing.html) - 48k - [Cached](#)

**Unique/Duplicates?**

# gTLD, languages & countries



## Advanced Search

### Find web pages that have...

all these words:

this exact wording or phrase:

one or more of these words:

### But don't show pages that have...

any of these unwanted words:

### Need more tools?

Results per page:

Language:

File type:

Search within a site or domain:   
(e.g. youtube.com, .edu)

### [Date, usage rights, numeric range, and more](#)

Date: (how recent the page is)

[Usage rights:](#)

Where your keywords show up:

Region:

[Advanced Search](#)  
[Preferences](#) Search the Web  Search Arabic pages[Web](#) [Images](#)

Results 1 - 10 of about 56,500,000 Arabic pages for site:com.

# Exalead languages & file types

The screenshot shows the Exalead search interface. At the top, there are navigation links for 'Web', 'Images', 'Wikipedia', 'Video', and 'More'. A search box contains 'site:uva.nl' and a 'Web Search' button. Below the search bar, a blue bar indicates 'Web Results 1-10 of about 117,989 for on uva.nl'. The main results area lists three items, each with a preview icon:

- Intelligent Systems Lab Amsterdam Research them**  
At ISLA we prefer to study hard scientific problems from real data: principle, the data and problems we work on are ...  
[www.science.uva.nl/research/isla](http://www.science.uva.nl/research/isla) - 7k - [Add to shortcuts](#)
- UvA Computer Museum**  
However the collection also includes tools from the pre-computer some word-processors and a small collection of home computer:  
[www.science.uva.nl/museum](http://www.science.uva.nl/museum) - 18 Sep 2007 - 4k - [Add to shortcuts](#)
- AIR**  
is now merged into the System and Network Engineering group f visitors in October 2007 last update 11 - 08 - 2006 fdijkstr@scien  
[www.science.uva.nl/research/air](http://www.science.uva.nl/research/air) - 09 Oct 2007 - 5k - [Add to shorts](#)

On the right side, there is a sidebar with two sections:

- Languages**
  - English (62%)
  - Dutch (35%)
  - German (1%)
  - French (0.1%)
  - Indonesian (0.1%)
  - Portuguese (0.1%)
  - Romanian (0.1%)
  - Italian (0.1%)
- File types**
  - Acrobat (.pdf) (53%)
  - Text (.txt) (33%)
  - Flash (.swf) (5%)
  - Word (.doc) (3%)
  - PowerPoint (.ppt) (2%)
  - Rich Text Format (.rtf) (1%)
  - Excel (.xls) (0.2%)
- Geographic location**
  - Europe (100%)
    - Netherlands

Web | [Images](#) | [Video](#) | [Local](#) | [Shopping](#) | [more](#) ▾

site:cern.ch originurlextension:pdf

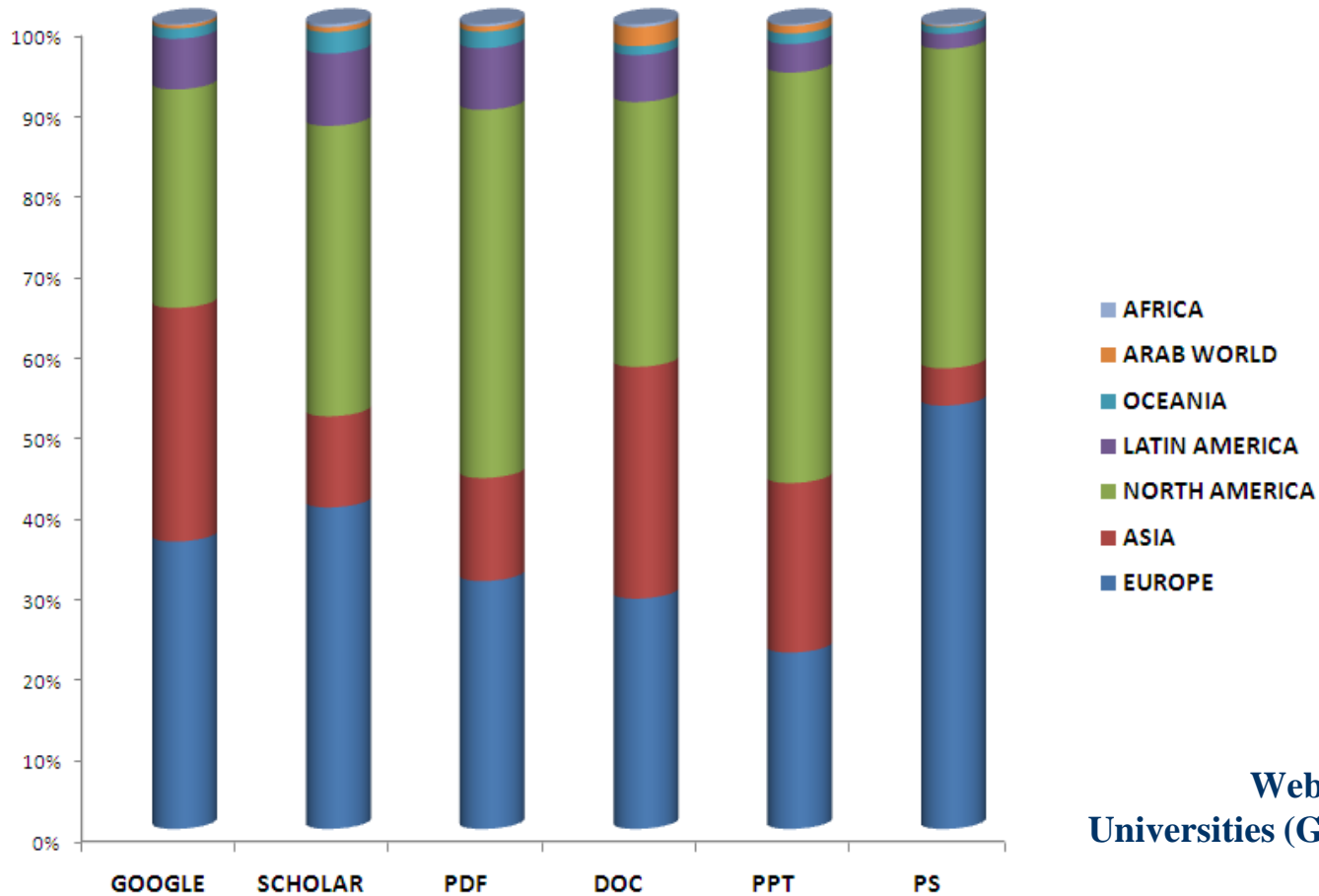
Search

YAHOO!

1-10 of 73,900 from cern.ch for originurlextension : pdf

1. [report of the task force on open access publishing in particle physics](#) (PDF)  
The Task Force is most grateful to Fred Friend whose intimate knowledge of the ... [cern.ch/Open-Access/20050916.html](#). 7. [http://indico.cern.ch ... library.cern.ch/OATaskForce\\_public.pdf](#) - 2458k - [View as html](#)
2. [Long Term Variation of the Circumference of the SPring-8 Storage Ring](#) (PDF)  
Tokai, Naka-gun, Ibaraki 319-1195, JAPAN. ... that in addition to the tidal deformation the circumference. possesses the long term variation. ... [accelconf.web.cern.ch/accelconf/e00/PAPERS/MOP5A04.pdf](#) - 663k - [View as html](#)
3. [www.datatag.org](#) (PDF)  
1. Technical Report DataTAG-2004-1. FP5/IST DataTAG ... [http://cern.ch/jpmf/ Yee-Ting Li. Department of Physics and Astronomy. University College London ... datatag.web.cern.ch/datatag/papers/tr-datatag-2004-1.pdf](#) - 680k - [View as html](#)
4. [An Application of the Theory of Permutations in Breaking the Enigma Cipher](#) (PDF)  
beginning some mathematical methods, mainly the elements of probability theory and ... in the twenties of our century, broadened considerably the field of ... [frode.home.cern.ch/frode/crypto/rew80.pdf](#) - 1626k - [View as html](#)
5. [-er /家/ -est /最/ -ian /家/ -ism /主义/ -ist /家/](#) (PDF)  
亚非新闻工作者协会/ AALPS /自动空运计划制定系统/ AAM /空对空导弹/ AARP /地址转换协议/ ABC /基本知识/英语字母表/火车时刻表/ ABM /反弹道飞弹/ AC /鉴别码/ ACL /存取控制表/访问控制包/ ADABAS /自适应数据库系统/ ADP /自动数据处理 ... [lshuai.web.cern.ch/lshuai/chinese/file/dict.pdf](#) - 3693k - [View as html](#)
6. [FISH and I](#) (PDF)  
I have been asked to speak today about some cryptographic work I was engaged in ... I was concerned mainly with a German machine-cipher known in Bletchley ... [frode.home.cern.ch/frode/crypto/tutte.pdf](#) - 63k - [View as html](#)

# Rich files & Scholar



Webpages from  
Universities (Google, January, 2009)

# Informal scholarly communication

- ◆ R&D web size can be estimated from the academic sub-domains contribution to webspace
  - Excluding administrative and other non-relevant pages and adding R&D sites under other domains the size of R&D could be over 10%
- ◆ Search engines index now rich files (*pdf, ps, ppt, doc, xls, rtf*) usually associated to formal and informal communication activities in the academic arena
  - Material for students
  - Researchers' personal home pages
    - Papers, conference presentations, drafts, raw data files
  - Department document archives
  - Electronic libraries (incl. thesis)
  - Subject repositories

## LARGEST UNIVERSITIES IN THE WEB

DOMAINS	Webpages	PDF	DOC	PPT	PS	XLS	Scholar
<i>kaist.ac.kr</i>	7420000	26300	4300	8330	1460	734	41500
<i>psu.edu</i>	7240000	89700	27300	8560	2030	<b>4110</b>	41700
<i>mit.edu</i>	5700000	<b>445000</b>	8770	7110	<b>17200</b>	2940	42900
<i>berkeley.edu</i>	4870000	212000	12900	14600	9680	2390	18900
<i>stanford.edu</i>	4770000	323000	13700	13000	13300	3780	41200
<i>harvard.edu</i>	4460000	97200	7370	3350	4420	1210	<b>378000</b>
<i>ualberta.ca</i>	4370000	52100	4680	3330	2760	1040	6310
<i>kyoto-u.ac.jp</i>	4280000	74500	2740	2000	1890	611	14600
<i>wisc.edu</i>	3930000	221000	21700	11500	6140	3310	18900
<i>washington.edu</i>	3530000	153000	20000	9720	5070	2620	9990
<i>uiuc.edu</i>	3350000	244000	19500	<b>28200</b>	6120	3530	17200
<i>arizona.edu</i>	3140000	93500	15800	4670	3150	2700	11300
<i>cam.ac.uk</i>	3100000	69900	3740	1940	4720	405	7560
<i>ucsc.edu</i>	2730000	67900	3070	2150	3310	1240	2500
<i>unc.edu</i>	2540000	53700	17800	8050	2050	2580	5480
<i>ntu.edu.tw</i>	2530000	50800	<b>33100</b>	10300	946	2140	36200
<i>ncsu.edu</i>	2430000	97900	9440	3880	1270	1300	9340
<i>ucsd.edu</i>	2400000	80300	5830	6900	4880	1420	7360
<i>utah.edu</i>	2400000	43600	4910	2440	2770	908	5780
<i>liu.se</i>	2370000	39600	3540	846	3460	462	6450
<b>14948 universities</b>	<b>634104299</b>	<b>27735767</b>	<b>6193873</b>	<b>1682537</b>	<b>832335</b>	<b>744946</b>	<b>4357181</b>

# (Google) Scholar



[Web](#) [Images](#) [Video](#) [News](#) [Maps](#) [more »](#)

site:cam.ac.uk

Search

[Advanced Scholar Search](#)

[Scholar Preferences](#)

[Scholar Help](#)

**Scholar** [All articles](#) - [Recent articles](#)

Results 1 - 10 of about 8,560

All Results

[P Mockapetris](#)

[S Deering](#)

[J Moy](#)

[J Holtzman](#)

[M Rose](#)

## [OSPF Version 2](#)

[J Moy - dspace.cam.ac.uk](#)

Network Working Group J. Moy Request for Comments: 1247 Proteon, Inc.  
Obsoletes: RFC 1131 July 1991 OSPF Version 2 Status of this Memo This RFC specifies an IAB standards track protocol for the ...

[Cited by 1727](#) - [Cached](#) - [Web Search](#)

## [Host extensions for IP multicasting](#)

[SE Deering - dspace.cam.ac.uk](#)

Network Working Group S. Deering Request for Comments: 1054 Stanford University  
Obsoletes: RFC 988 May 1988 Host Extensions for IP Multicasting 1. STATUS OF THIS MEMO This memo specifies the extensions required of a host implementation of ...

[Cited by 1521](#) - [Cached](#) - [Web Search](#)

## [Domain names-concepts and facilities](#)

[PV Mockapetris - dspace.cam.ac.uk](#)

Network Working Group P. Mockapetris Request for Comments: 1034 ISI Obsoletes: RFCs 882, 883, 973 November 1987 DOMAIN NAMES - CONCEPTS AND FACILITIES 1. STATUS OF THIS MEMO This RFC is an introduction to the Domain Name System (DNS), and ...

[Cited by 1169](#) - [Cached](#) - [Web Search](#)

# Publish or Perish

## Query

Author(s):

Publication:

All of the words:

Any of the words:

None of the words:

The phrase:

Year of publication between:  and:

Title words only

- Biology, Life Sciences, Environmental Science
- Business, Administration, Finance, Economics
- Chemistry and Materials Science
- Engineering, Computer Science, Mathematics
- Medicine, Pharmacology, Veterinary Science
- Physics, Astronomy, Planetary Science
- Social Sciences, Arts, Humanities

Lookup  
Lookup Direct  
Help

## Results

Papers: 820 Cites/paper: 2.80 h-index: 21 AWCRC: 362.71  
 Citations: 2293 Cites/author: 1258.11 g-index: 28 AW-index: 19.04  
 Years: 9 Papers/author: 520.26 hc-index: 15 AWCRCpA: 197.30  
 Cites/year: 254.78 Authors/paper: 2.25 h1-index: 7.23  
 h1\_norm: 14

Copy statistics  
Copy results  
Check all  
Check selection  
Uncheck all  
Uncheck 0 cites  
Uncheck selection  
Help

Cites	Per year	Rank	Authors	Title	Year	Publication	Publisher
<input checked="" type="checkbox"/>	60	8.57	1	D Nualart, A Rascanu	2002	Differential equations driven by frac... Collect. Math	imub.ub.es
<input checked="" type="checkbox"/>	43	4.78	2	P Camps, R El Acha...	2000	New tacrine-huperzine A hybrids (hu... J. Med. Chem	ub.es
<input checked="" type="checkbox"/>	38	5.43	5	H Dvir, DM Wong, ...	2002	3D structure of Torpedo californica a... Biochemistry	ub.es
<input checked="" type="checkbox"/>	38	5.43	4	J Arbiol, J Cerda, G ...	2002	Effects of Nb doping on the TiO 2 an... Journal of Applied Physics	nun97.el.ub.es
<input checked="" type="checkbox"/>	37	4.11	3	A Crisanti, F Ritort, ...	2000	Inherent structures and nonequilibri... The Journal of Chemical Phy...	ffn.ub.es
<input checked="" type="checkbox"/>	36	4.50	6	MA Nielsen, G Vidal	2001	Majorization and the interconversion... Quantum Information and C...	sophia.ecm.ub.es
<input checked="" type="checkbox"/>	36	5.14	44	U BECK	2002	Libertad o capitalismo: el incierto fut... Madrid: El País, Opinión-17.(...	ub.es
<input checked="" type="checkbox"/>	35	5.00	13	SY Liu, JM Girart, A ...	2002	Formic Acid in Orion KL from BIMA 1 ... ApJ	am.ub.es
<input checked="" type="checkbox"/>	35	4.38	7	G Dinca, P Jebelean...	2001	Variational and topological methods f... Portugaliae Mathematica	mat.ub.es
<input checked="" type="checkbox"/>	35	5.83	8	L Bértola, JG William...	2003	Globalization in Latin America Before ...	ub.es
<input checked="" type="checkbox"/>	31	6.20	9	NJ Harmer, L Pelleg...	2004	The crystal structure of fibroblast gr... Biochemistry	mmb.pcb.ub.es
<input checked="" type="checkbox"/>	28	5.60	11	M Rodriguez, A Sto...	2004	The effect of private insurance acce... Health Economics	eco.ub.es
<input checked="" type="checkbox"/>	27	3.38	14	L Coutin, D Nualart,...	2001	Tanaka formula for the fractional Br... Stochastic Proc. Applic	mat.ub.es
<input checked="" type="checkbox"/>	27	3.00	10	P Gorostiza, M Kula...	2000	Charge exchange processes during t... Journal of the Electrochemi...	qf.ub.es
<input checked="" type="checkbox"/>	25	2.78	12	J5 Laberg, TO Vorren	2000	The Trifnadupet Slide, offshore Nor... Marine Geology	geomar.geo.ub.es
<input checked="" type="checkbox"/>	24	6.00	15	L Crespo, G Sandlm...	2005	Peptide and amide bond-containing ... Chem Rev	pcb.ub.es
<input checked="" type="checkbox"/>	23	4.60	23	JP Gee	2004	Learning by Design: Games as Learni... Interactive Educational Multi...	ub.es
<input checked="" type="checkbox"/>	22	3.14	16	D Nualart, Y Ouknine	2002	Regularization of differential equatio... Stochastic Processes and th...	mat.ub.es
<input checked="" type="checkbox"/>	21	3.00	19	B Fernandez-Manjo...	2002	Creating cost-effective adaptive ed... Interactive Educational Multi...	ub.es
<input checked="" type="checkbox"/>	21	2.63	20	A Jorba, M Zou	2001	A software package for the numeric... Preprint	zeus.maia.ub.es
<input checked="" type="checkbox"/>	21	2.63	18	JM Spector	2001	An Overview of Progress and Proble... Interactive Educational Multi...	ub.es
<input checked="" type="checkbox"/>	20	2.86	21	P Cironi, I Manzana...	2002	Solid-Phase Total Synthesis of the P... Tetrahedron Lett	pcb.ub.es
<input checked="" type="checkbox"/>	20	3.33	17	G Massana	2003	Parahippocampal Gray Matter Densit... American Journal of Psychiatry	ub.es
<input checked="" type="checkbox"/>	18	2.00	29	R Sayos	2000	Theoretical investigation of the eight... The Journal of Chemical Phy...	qf.ub.es
<input checked="" type="checkbox"/>	17	3.40	30	M Fontes, X Verdag...	2004	2-Piperidino-1, 1, 2-Triphenylethanol... J. Org. Chem	pcb.ub.es
<input checked="" type="checkbox"/>	17	1.89	22	X Verdaguer, A Moy...	2000	A New Chiral Bidentate (P, S) Ligand... Tetrahedron: Asymmetry	pcb.ub.es
<input checked="" type="checkbox"/>	17	1.89	24	GL Karakostas, PC ...	2000	POSITIVE SOLUTIONS OF A BOUND... Electronic Journal of Differe...	mat.ub.es
<input checked="" type="checkbox"/>	16	2.29	26	M Acet, E Duman, E...	2002	Coexisting ferro-and antiferromagne... Journal of Applied Physics	ecm.ub.es
<input checked="" type="checkbox"/>	16	2.29	27	G Gomez, JM Mond...	2002	Refined Fourier analysis: procedures... Preprint	zeus.maia.ub.es
<input checked="" type="checkbox"/>	16	2.67	31	JJ Conde-Ruiz, P Pr...	2003	What Social Security: Beveridgean O... ub.es	ub.es
<input checked="" type="checkbox"/>	16	2.29	25	R Laugesen	2002	Translational averaging for complete... Collect. Math	imub.ub.es
<input checked="" type="checkbox"/>	16	2.00	28	J Lesperance	2001	Gorenstein liaison of some curves in ... Collectanea Mathematica	imub.ub.es
<input checked="" type="checkbox"/>	15	1.88	104	I Ramonet	2001	'El consenso de Porto Alegre' El País	ub.es

# II

Web indicators based Rankings  
as a tool for the evaluation of  
academic and scientific  
institutions

# Building a model

- ◆ Inputs
  - Candidate web editors
  - Funding
- ◆ Outputs
  - Web size
  - Rich files
  - Media files (?)
  - Academic papers
- ◆ Visibility (Impact)
  - Link visibility
- ◆ Usage (Popularity)
  - Visits and visitors
  - Downloads

# In the academia (I)

- ◆ Candidate web editors
  - Central organization personnel: directorship, administration, libraries, press offices, external (inc. International) relationships, press offices, printing offices
  - Faculty staff: professors and researchers
  - Students: postgraduates, undergraduates
  - Email accounts
- ◆ Funding
  - Total budget
  - Budget devoted to ICT
  - Servers and Web sites
  - Research projects

# In the academia (II)

- ◆ Web size
  - Number of pages
  - Number of pages indexed by search engines
  - Subdomains (and autonomous websites: blogs, wikis, ...)
- ◆ Rich files
  - Number of files by type
  - Files in repositories/subdomains
- ◆ Media files
  - Number of files by type
- ◆ Academic papers
  - Personal pages
  - Repositories
  - Databases

# In the academia (III)

- ◆ Link visibility
  - External inlinks
  - InLinks diversity
  - PageRank (weighted links)
  - Self-links
- ◆ Visits and visitors
  - Targets (directories and files popularity)
  - Geographical/institutional distribution
  - Temporal distribution
  - Informetrics of referrers/referrals
- ◆ Downloads
  - Bibliometrics

# Analysis

- ◆ Institutional
  - Research universities
  - Comprehensive universities
  - Specialised universities
  - Open universities
  - Research institutes
- ◆ Geographical
  - World
  - Regional, national, local
- ◆ Disciplines
  - Technological vs. Non-technological
  - Contribution of social sciences and humanities
  - Knowledge transfer related topics

# Ranking Web of World Universities

January 09

## Top 4000 Universities

[First](#) | [Previous](#) | [Next](#) | [Last](#) | Universities 1 to 50 of 4000

WORLD RANK	UNIVERSITY	COUNTRY	POSITION			
			SIZE	VISIBILITY	RICH FILES	SCHOLAR
1	Massachusetts Institute of Technology		1	3	2	6
2	Stanford University		2	2	3	12
3	Harvard University ***		3	1	17	1
4	University of California Berkeley		6	4	5	24
5	Cornell University		4	5	8	37
6	University of Michigan		10	6	15	22
7	California Institute of Technology ***		8	8	21	17
8	University of Minnesota		9	16	4	19
9	University of Illinois Urbana Champaign *		14	10	6	38
10	University of Texas Austin		11	9	10	45
11	University of Wisconsin Madison		5	13	9	47
12	University of Washington		16	7	7	63
13	University of Pennsylvania		13	12	33	27
14	Pennsylvania State University ***		18	21	22	18
15	Carnegie Mellon University		7	25	1	51
16	Texas A&M University		20	31	14	11
17	Columbia University New York		22	15	19	58
18	University of California Los Angeles		15	18	23	70
19	University of Maryland		30	30	16	36
20	University of Chicago		51	14	75	3
21	Purdue University		21	29	12	73
22	University of Florida		19	34	11	75

## Top 500 European Universities

1-100 | [101-200](#) | [201-300](#) | [301-400](#) | [401-500](#) | Universities 1 to 100 of 500

RANKING	UNIVERSITY	COUNTRY	POSITION				
			EUROPE	WORLD	SIZE	VISIB.	RICH FILES
1	University of Cambridge		28	29	23	55	110
2	Swiss Federal Institute of Technology ETH Zurich		40	35	67	52	26
3	University of Oxford		42	62	27	59	127
4	University of Helsinki***		58	45	96	80	49
5	University of Oslo		60	55	99	30	88
6	University of Edinburgh		65	137	78	54	93
7	University College London		68	209	55	56	114
8	Utrecht University		71	121	122	70	32
9	University of Manchester ***		82	147	79	123	169
10	Royal Institute of Technology		84	79	133	69	138
11	Freie Universitat Berlin		85	105	109	160	96
12	Universitat Wien		90	99	107	50	241
13	Università di Bologna		93	110	116	121	145
14	Université Catholique de Louvain *		95	103	85	247	155
15	Université de Geneve		96	168	97	91	185
16	University of Glasgow		99	182	65	156	249
17	Norwegian University of Science & Technology		101	53	200	92	39
18	University of Amsterdam		103	91	152	83	164
19	Uppsala University		104	61	175	81	133
20	École Polytechnique Federale de Lausanne		105	94	173	113	74
21	Humboldt Universitat zu Berlin		106	71	180	111	103

# Rankings

CRITERIA	WEBOMETRICS	ARWU	HEEACT	QS-THES
	Spain (research lab)	China (university)	Taiwan (gov agency)	United Kingdom (firm)
Univs analysed	17000	3000	3500	2000
Univs ranked	6000	500	500	500+
Teaching		Alumni Nobel 10%		Students/Staff 20%
Internatio- nalisation				% Foreign Students 5% % Foreign Professors 5%
Size	Webpages 20%	University 10%		
Research	Documents 15% Papers GS 15%	Nature&Science 20% Papers S&SSCI 20%	Papers S&SSCI 20%	
Impact	Links 50%	Highly Cited Researchers 20%	Highly Cited Res 50% Cit High IF Journals 30% Citations S&SSCI 30%	Citations Scopus 20%
Prestige		Faculty Nobel 20%		Academic Survey 40% Employers Survey 10%

CRITERIA	INDICATOR		SOURCES
SIZE	S	NUMBER OF PAGES	Google, Yahoo, Live & Exalead
	R	NUMBER OF RICH FILES	Google
	Sc	NUMBER OF PAPERS	Google Scholar
		NUMBER OF PAGES IN ENGLISH	
VISIBILITY	V	NUMBER OF EXTERNAL INLINKS	Yahoo, Exalead, (Live)
LUMINOSITY		NUMBER OF EXTERNAL OUTLINKS	
SUBDOMAINS		NUMBER OF SUBDOMAINS	
POPULARITY		NUMBER OF VISITS	

### Normalization

$$N_a = \frac{\log(n_a + 1)}{\log(\max(n_i) + 1)}$$

N: search engine (Google, Yahoo, Live, Exalead)  
a: web domain

### Median method (Size)

$$S_a = \frac{1}{2} * ((G_a + Y_a + L_a + E_a) - \max(G_a, Y_a, L_a, E_a) - \min(G_a, Y_a, L_a, E_a))$$

### Rich files

$$R_a = PDF_a + DOC_a + PPT_a + Ps_a$$

### Ranking

$$S_a \rightarrow R_a(S)$$

$$R_a \rightarrow R_a(R)$$

$$Sc_a \rightarrow R_a(Sc)$$

$$V_a \rightarrow R_a(V) \quad 1:1$$

$$WR = (50\% * V) + (20\% * S + 15\% * R + 15\% * Sc)$$

# Digital Divide

REGION / countries		Top100	Top200	Top500	Top1000	Total
<b>NORTH AMERICA</b>	7	<b>78</b>	<b>121</b>	206	392	<b>3552</b>
USA		70	104	179	354	<b>3343</b>
<b>EUROPE</b>	54	<b>16</b>	<b>60</b>	223	412	<b>4531</b>
United Kingdom		6	12	35	68	<b>235</b>
Germany		1	17	49	67	<b>384</b>
Italy		1	1	13	35	<b>200</b>
France			1	16	47	<b>603</b>
Spain			4	27	42	<b>218</b>
<b>ASIA</b>	33	3	10	40	118	<b>4303</b>
Japan		2	3	10	38	<b>670</b>
China & HK			3	10	22	<b>1156</b>
Taiwan		1	2	10	23	<b>158</b>
Korea			1	4	12	<b>402</b>
<b>LATIN AMERICA</b>	35	2	3	13	36	<b>2891</b>
Brazil		1	2	9	16	<b>1529</b>
Mexico		1	1	2	6	<b>354</b>
<b>OCEANIA</b>	12	1	6	15	35	<b>131</b>
Australia		1	6	13	29	<b>79</b>
<b>ARAB WORLD</b>	22			2	2	<b>496</b>
<b>AFRICA</b>	37			1	5	<b>321</b>
<b>WORLD</b>	<b>200</b>					<b>16225</b>

# Biases?

Country	WEBR09	ARWU08	HEEACT08	QSTHES08
USA	<b>104</b>	90	91	<b>58</b>
United Kingdom	12	22	19	<b>30</b>
Germany	17	14*	18	<b>11</b>
Canada	<b>17</b>	<b>6</b>	9	12
Netherlands	<b>6</b>	9	9	10
Japan	<b>3</b>	9	9	10
Australia	6	6	6	<b>9</b>
Switzerland	<b>3</b>	6	7	7
France	<b>1</b>	7	4	4
Sweden	4	4	4	4
Italy	1	<b>5</b>	<b>7</b>	1
Israel	1	4	3	3
Belgium	<b>1</b>	4	2	5
Denmark	<b>1</b>	3	2	3
China & HK	3	0	3	<b>10</b>
Korea	1	1	1	<b>3</b>
Singapore	1	1	1	2
Taiwan	2	1	1	1
India	0	0	0	<b>2</b>
Thailand	0	0	0	1

# Summary (I)

- ◆ Cybermetrics/Webometrics is not only bibliometrics of the Web
  - ◆ But electronic journals, repositories and other electronic scholar communication objects could be analyzed using bibliometrics methodology
- ◆ Web indicators could be useful alone or complementing scientometric ones
  - ◆ New intermediaries, the search engines, should receive special attention
  - ◆ Link analysis (citation + “sitation” + other motivations) need to be developed in depth
  - ◆ Web “big numbers” is a great advantage

# Summary (II)

## ◆ Some results

- ◆ Institutional web domains is a natural unit for Webometrics
  - ◆ Size and visibility indicators can be derived from search engines
- ◆ Open access initiatives are changing the way researchers communicate their results
  - ◆ Rich file formats and web-citation databases (Scholar) are becoming a “third” source (WoS, Scopus)
- ◆ Rankings assuming the role of impact factor in 70s & 80s
- ◆ Positioning, surprisingly, a powerful tool for web publication
- ◆ Log file analysis and popularity, a new hope for the “circulation”-based studies?



---

Questions?

---

Thank you!

Isidro F. Aguillo  
[isidro.aguillo@cchs.csic.es](mailto:isidro.aguillo@cchs.csic.es)