> Homepage  > The Journal  > Issues Contents  > Vol. 16 (2012)  > Paper 3        June 6, 2012

**The Journal**
Cybermetrics News
Editorial Board
Guide for Authors
Issues Contents
**The Seminars**
**The Source**
Scientometrics
Tools
R&D Policy & Resources

**VOLUME 16 (2012): ISSUE 1. PAPER 3**

## Graphic, multimedia, and blog content presence in the Spanish academic web-space

**Enrique Orduña-Malea**

Universitat Politècnica de València
Valencia, **Spain**
E-mail: enorma@upv.es

### Abstract

The number of pages in a website is an indicator (related to its activity) widely used in cybermetric analysis. This indicator can be disaggregated by type of content and file type. In this sense, a gap in the literature about the treatment and quantitative analysis of multimedia files, graphics and type blog is detected, and particularly in their presence and distribution in the academic environment. This paper proposes a diachronic analysis in 2010 of media and graphic files count, and blog-like content for all websites which conforms the Spanish university space. Among the key findings, a very high percentage of blog-like content and image files are detected, which contrasts with the very low figures obtained for multimedia files. Otherwise, diverse limitations in image searchers used are found (coverage, variations between samples, instability and discrepancies between the calculation of global and file format counts), which call for a careful interpretation of the raw results obtained. Finally, a correlation between *Bing images* and *Google images* higher than expected (limited by a small set of URLs), and a sharp decrease on Bing coverage during the study period is obtained.

### Keywords

Spain; Universities; Higher education institutions; Spanish university system; Graphic files; Multimedia files; Blog contents; Image search, Webometrics; Cybermetrics; Quantitative webometrics; World wide web; www; Bing; Google, Visibility

### 1. Introduction

The main indicators used for cybermetric purposes can be broadly classified into the following categories (Aguillo, 2009):

- Indicators related to the activity: for example, site size.
- Indicators related to the impact: for example, external inlinks or textual mentions.
- Indicators related to the usage: for example, the number of downloads.

Within the topic area dedicated to the study of academic web-spaces (universities, and other higher education institutions), the latter type of indicators is rarely used due to accessibility problems of certain indicators (ie, web traffic); for that reason, indicators related with size and visibility are most commonly used (Aguillo et al., 2006; Orduña-Malea et al., 2010).

Focusing on size indicators, we can identify mainly the following two types (Aguillo, 2000):

- File size: the number of bytes in a file, or online folder/subsite/site.
- Site size or document count: the number of pages (files) in an online folder/subsite/site.

The site size can be classified into global count or specific count (that is, a measure according to the type of nature/format/file that a web unit publishes). Some important specific counts according to the file are: office files (such as DOC, PPT or PDF), graphic files (such as JPG, PNG or BMP), web files (such as ASP, PHP or HTML, etc.), and multimedia (such as MPG, WMV, MOV, etc.). These files can contain contents in one or more different formats (such as blog-like content, books, journals, papers, etc.), which can have one or different nature or purposes (academic, teaching, informative, etc.).

Of all these indicators, the most commonly treated are the global count (Aguillo, Ortega & Fernandez, 2008), academic specific count (Orduna-Malea et al., 2009), and within specific count, the so-called rich files. The latter are important because many of them are entire papers or other scientific documents (Aguillo, 2009). So they are good indicators of academic published information (Kousha & Thelwall, 2008; Kousha, Thelwall & Rezaie, 2010).

However, within the set of existing files, other formats are poorly treated from a quantitative point of view (but important due to the amount generated), although they do not express directly the functional activities of the university. These include, apart from purely web files (HTML, PHP, ASP, etc.), blog content type, multimedia files, and graphics files.

The value of the blog content type as a communication tool for social institutions (including universities) is well established today (Goodfellow & Graham, 2007), and the existence of blog search engines (Thelwall & Hasler, 2007) has allowed the quantitative analysis from the discipline of cybermetrics (Thelwall, 2009), although primarily aimed at studying the spread of ideas (Thelwall & Price, 2006), seeking points of views (Thelwall, 2007), interests of certain groups (Thelwall & Prabowo, 2007), analysis of general issues (called "blog issue analysis": Thelwall & Wilkinson, 2010) or even emotions (Thelwall et al., 2010).

The analysis of multimedia files from a quantitative point of view is still very uncommon (almost nonexistent in academic environments). The few identified studies focus on the analysis of user-generated content on *Youtube*-like platforms (Cha et al., 2009; Thelwall, South & Vis, 2012; Kousha, Thelwall & Abdoli, forthcoming). The availability of access to Youtube's API suggests that this type of analysis can be further developed in the future.

In the case of graphic files and images also abound analysis about user-generated content (tagging) on platforms, in this case predominantly Flickr (among others, Sigurbjörnsson & Zwol, 2008; Angus, Thelwall & Stuart, 2008; Angus & Thelwall, 2010).

Moreover, the fact that the users' access to graphic files is recorded in the log files has opened up new working lines, among which are the work of Chen (2001), and Choi & Rasmussen (2003), centered in the analysis of image queries.

Another interesting working line is such about the persistence of digital objects on the Web (Koehler, 1999), where among the items discussed are the image files, and where already indicated the large volume of existing graphic files. In line with this topic, Ortega, Aguillo & Prieto (2006) found that graphic and multimedia files had an important growth rate (10.50%, and 6.43% respectively, between 1997 and 2004, from 738 selected websites all over the world), and also a highly vanishing rate (80.34% of image elements).

files, such as video, audio, and graphic files.

Also highlights another working line aimed at finding images on the Web, which are worth mentioning the work carried out on Excite (Goodrum & Spink, 2001) and other special search engines like TinEye[1] (Kousha, Thelwall & Rezaie, 2012). Finally, should also be remarked research focused on comparative analysis between text and image searches on the Web (Pu, 2005; 2008), and the general search of digital images (Jansen, 2008).

However, the classic studies of cybermetrics applied to the quantification of universities' web performance have not paid much attention about graphic files, in large part because the motivations for the creation and/or reuse of these images respond to very different purposes, and they don't reflect only the core activities of these institutions (Kousha, Thelwall & Rezaie, 2012). Notwithstanding, there are some recent studies that analyze the use and impact of the images as a resource in academic activities (Angus, Thelwall & Stuart, 2010; Angus & Stuart, 2010), although none of these works take the academic website as the unit of measure.

Despite the quantification of such files is not as direct indicator of the activities of the university as rich files, its volume is very high, and since the global count it is used in universities' web measurements, their nature and influence should be studied more precisely.

As regards the Spanish area, it deserves special attention the work of Alonso, Figuerola & Zazo (2004), and Pinto et al. (2004), the only existing studies where graphic files are quantified within the Spanish academic web-space. However, the sample used (only a part of the Spanish university system), method of analysis (web crawler rather than a commercial search engine), and the obsolescence of their results (conducted in 2004) confirm the need to update the results.

For all these reasons, the main objective of this work is the quantification of the blog-like content, graphic and multimedia files within the Spanish university system, for which the following specific objectives are proposed:

- To quantify the count of graphic and multimedia files and blog-like content within the Spanish university web-space over a full year (2010).
- To discover the proportion of these files within the number of pages in the websites of Spanish universities for the same period of study.
- To identify and analyze the differences between special search engines used to measure over time, and the possible limitations and inconsistencies of these.

## 2. Method

### 2.1. Data gathering

The analysis is applied to the Spanish university system, formed in 2010 by 76 universities, both public and private. The list of universities and associated URLs were obtained from the Ministry of Education and the Conference of Rectors of Spanish Universities (CRUE)[2]. In addition to the official URLs (those indicated in the official sources consulted), the existence of alias and "alternative" domains, at various universities was detected.

Regarding alias domains (URLs that share the same second level domain, but have different top level domain), it was found manually at each university the existence of the following domains: .CAT, .COM, .EDU, .ES, .NET, .ORG.

As regards the "alternative" domains (valid web domains with different first and/or second level domain respect to the official, and not necessarily redirected to it), they were searched through *Yahoo! Site Explorer*, as well as to consult the universities' institutional information on their website.

### 2.2. Measurement of the sample (I): indicators and sources

All indicators, scope definitions, sources, and commands utilized -with the exception of rich files, widely reported in the literature (Aguillo et al., 2006)- are showed below (table 1), where "domain.tld" should be substituted by each URL under study.

Table 1. Indicators, scope, sources, and commands

| INDICATOR | SCOPE | SOURCE | COMMAND |
|---|---|---|---|
| Global blog count | Number of retrieved documents that have been published in a blog type web platform | *Google Blogs* | blogURL:domain.tld |
| Global graphic file count | Number of retrieved files in any graphic format | *Google images* *Bing images* | site:domain.tld |
| Global multimedia count | Number of retrieved files in any multimedia format | *Bing videos* *Google videos* | site:domain.tld |
| Global count | Total number of retrieved files, without any restriction | *Bing* | site:domain.tld |
| Graphic file count | Number of retrieved files with a specific graphic file | *Google images* | site:domain.tld filetype:jpg site:domain.tld filetype:gif site:domain.tld filetype:bmp site:domain.tld filetype:png |
| Rich file count | Number of retrieved files with a specific office suite file | *Google* | site:domain.tld filetype:pdf site:domain.tld filetype:doc site:domain.tld filetype:ppt site:domain.tld filetype:xls site:domain.tld filetype:ps |

File formats provided by new versions of the *Microsoft Office* suite, such as DOCX and PPTX, were not considered due to their low representativeness in the period of measurement (2010), but should be considered in future works because they are estimated to have a larger growth rate.

In all sources, through the advanced settings of the browser, the following preliminary operations were performed:

- Deactivation of the parental control filter.
- Configuration of the number of maximum results per screen (100 in the case of *Google*, and 50 in the case of *Bing*).

**Bing images**
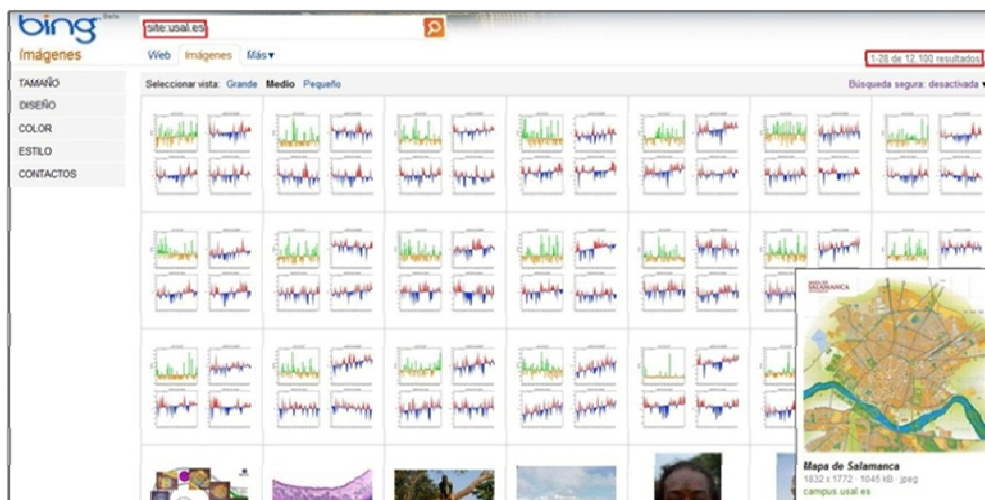Retrieve from: **http://www.bing.com/images** on 01-05-2011.

Figure 1. Example of global graphic count query on Bing

**Bing videos**
Retrieved from http://www.bing.com/videos
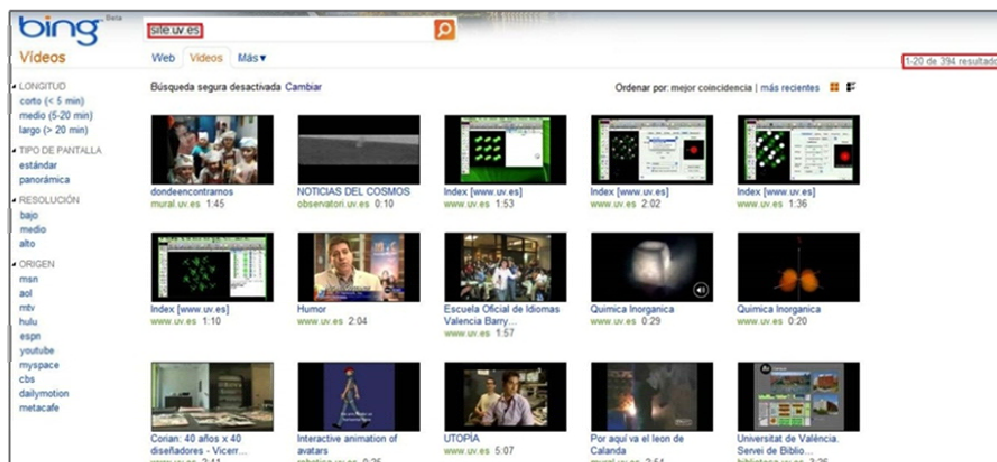


Figure 2. Example of global multimedia count query on Bing videos

**Google Blogs**
Retrieved from http://blogsearch.google.com/



Figure 3. Example of global blog count query on Google blogs

**Google images**
Retrieved from http://images.google.com/

Figure 4 illustrates the query process of graphic count on Google. As it can be seen, although the query can be set directly as shown in table 1, Google reconstructs the query in a general search ("site:domain.tld"), and then select the appropriate file from the sidebar menu box.

**Figure 4. Example of file graphic count (PNG) query on Google image**

*Google videos*
Retrieved from http://video.google.com/

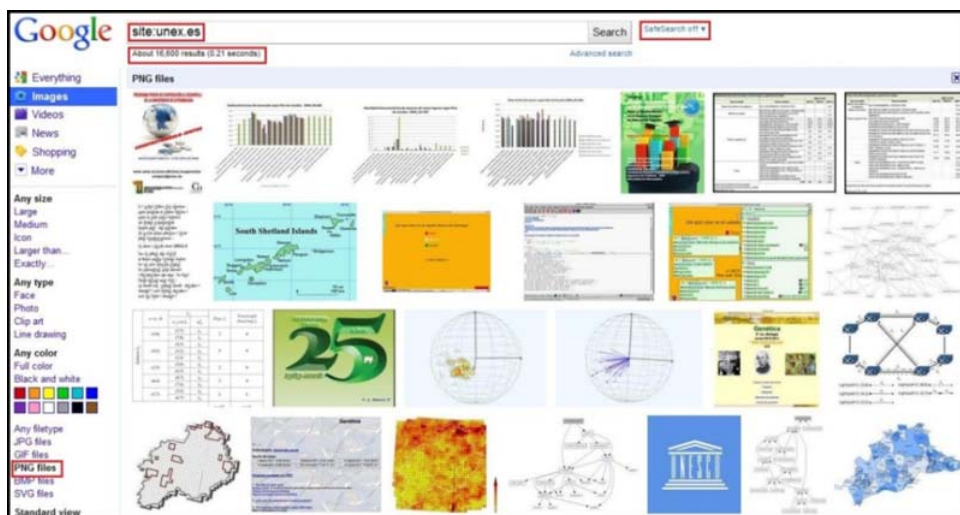Google announces on April 29th, 2011 that the Google videos service would stop working permanently on May 13th, 2011 due to competition with Youtube, also owned by Google[3]. Although all along 2010 still can be queried, it will affect the results, as discussed later (retrieved on 03/27/2012).
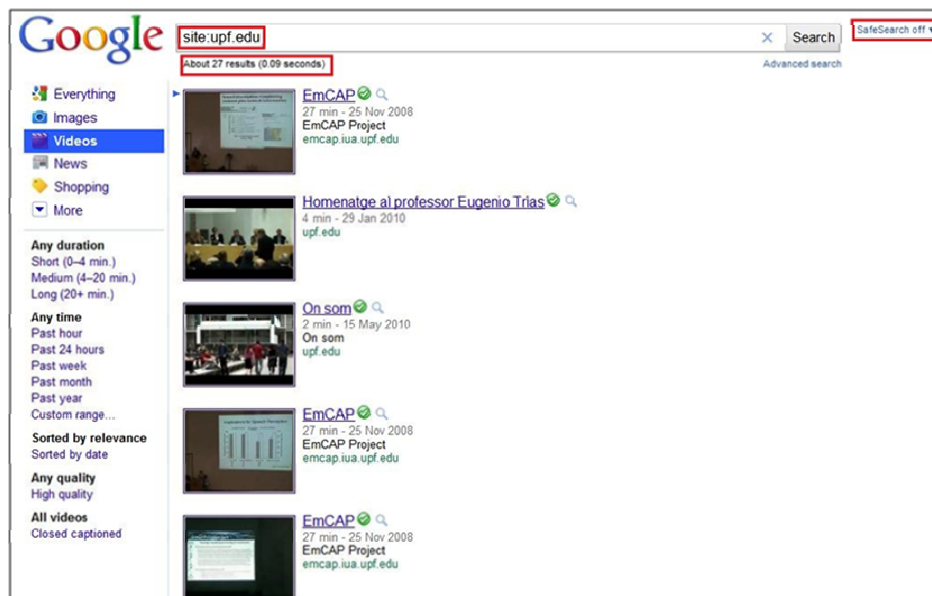


**Figure 5. Example of global multimedia count query on Google videos**

## 2.3. Measurement of the sample (III): data capture

The date of each measurement of the data simple is as follows:

- Sample 1: from 22nd to 31st March, 2010.

- Sample 2: from 21st to 30th June, 2010.

- Sample 3: from 20th to 29th September, 2010.

- Sample 4: from 20th to 31st December, 2010.

## 2.4. Analysis of the sample

Since the joint display of the raw data of all URLs is not adequate, due to the orders of magnitude (both among themselves and between the different search engines) are very different, it was decided to normalize count from 0 to 100, by a process of transformation (Rocki, 2005), with the aim of working subsequently with the called "mean relative representation factor in count" (Rc) (Orduña-Malea et al., 2010).

To this end, the sum of website size (in any of the indicators considered) obtained from the URLs of all Spanish universities over a whole month (accumulated count) is considered equal to 100, and the value of each URL is calculated proportionately:

$$v_{cn} = \frac{x_{cn}}{\sum_{i=1}^{N} x_{cn}} \cdot 100$$

[equation 1]

$v_{cn}$ = Normalized value obtained in count (c) for an URL (n).
$X_{cn}$ = Raw valued obtained in count (c) for an URL (n).
N = Set of URLs considered.

After normalizing results, total quantitative changes are not retrieved, but a size percentage proportional to the total size obtained by all the universities in a particular search engine every month measurement (4 shots in this case), a concept called "relative representation".

Then, the average of $v_{cn}$ is calculated monthly, obtaining a value, also between 0 and 100, whose name is "mean relative representation factor in count (Rc)". This factor can be calculated for any set of selected sites and any period of time (Orduna-Malea et al, 2010).

$$R_c = \frac{\sum_{i=1}^{M} v_{cn}}{M}$$

[equation 2]

Where M is the number of months analyzed (in this case 4 shots, for March, June, September and December 2010).

Finally, in order to calculate the growth rate of web domains along the period, the compound interest formula was used:

$$A = P\left(1 + \frac{r}{n}\right)^{n \cdot t}$$

[equation 3]

A= Accumulated account after n years.
P= Principal amount.
r= Annual rate of interest.
n= Number of times the interest is compounded per year.
T= Number of years.

## 3. Results

The results are divided into global results (accumulated count of each web domain in each shot), and results according to the university web domains.

### 3.1. Accumulated count

The global results for each search engine used (*Google* and *Bing*) are shown below:

*a) Google*

The inability to obtain accurate global count from *Google* prevents a full comparison (and hence, percentages) with the specific files considered (blog content, multimedia files, and rich files). In any case, figure 6 shows the evolution during 2010 of all counts measured by *Google*, including rich files (summation of PDF, DOC, PPT, XLS, and PS files).
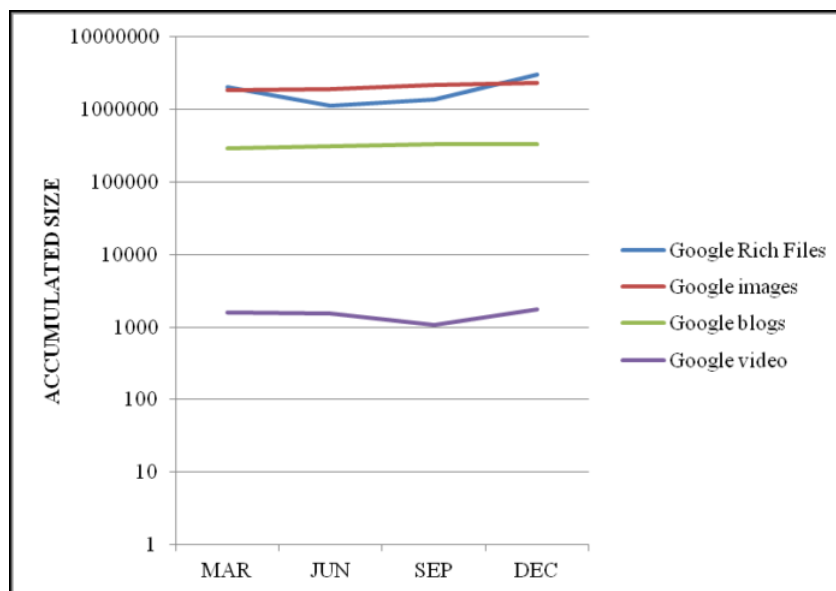


**Figure 6**. Comparison of count according to different *Google* searches

The most important files for all URLs that comprises the Spanish academic web space can be clearly identified: rich files (with an important drop in June), and the global graphic files, followed by blog contents, demonstrating the importance of such files in calculating the total number of pages in an online academic site.

The high values obtained both from blogs and graphic files aims to consider the percentage of this type of contents respect to the other considered files. Figure 7 displays this distribution (December shot), showing that 41% of all *Google* files considered come from graphic files, whereas 6% come from blogs.
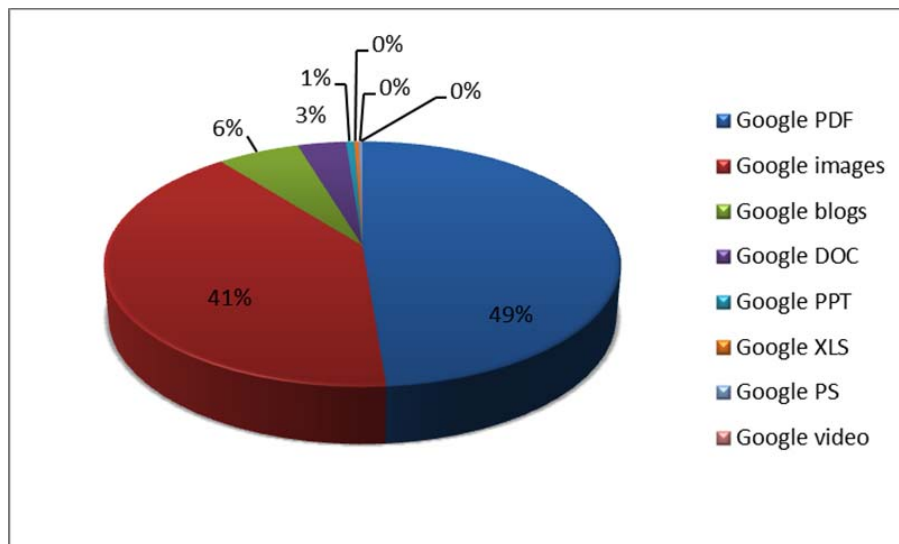
**Figure 7. Distribution of different files retrieved by *Google*:**
**rich, graphic, multimedia, and blogs (December 2010)**

As regards graphic formats, table 2 summarizes the full values obtained at each sample, which notes the widespread use of file JPG (56.10% of all considered files as for December sample), followed by GIF, and PNG. The BMP format usage is minority.

**Table 2. Raw graphic file count evolution, and annual nominal interest rate (r)**

| FORMAT | MAR | JUN | SEP | DEC | DEC (%) | r (%) |
|---|---|---|---|---|---|---|
| JPG | 1,149,438 | 1,093,254 | 1,247,131 | 1,342,190 | 56.10 | 0.16 |
| GIF | 608,698 | 716,538 | 798,779 | 790,918 | 33.06 | 0.27 |
| BMP | 3,861 | 5,132 | 5,606 | 5,500 | 0.23 | 0.37 |
| PNG | 318,845 | 203,494 | 238,248 | 253,990 | 10.62 | -0.22 |
| Total | 2,080,842 | 2,018,418 | 2,289,764 | 2,392,598 | 100% | 0.14 |
| Global | 1,881,481 | 1,926,775 | 2,164,129 | 2,305,847 | | 0.21 |

All formats present a positive statistical range throughout 2010, except PNG, which suffer a very substantial fall in the June data (also detected in JPG); although the data grows since then, it didn't reach the March levels again. This issue is reflected in the low growth rate (using compound interest formula) obtained.

The row "Total" shows the summation of all 4 types of formats considered, whereas "Global" row shows the global graphic count. This data shows the first inconsistencies: global count is lower than total count in all samples, although the figures keep the same order of magnitude.

*b) Bing*

Table 3 compares the global count with the graphic and multimedia files, showing the percentage of the specific counts respect to the global one for each shot.

**Table 3. Global, graphic, and multimedia file count for Bing**

| SAMPLE | BING | BING IMAGES | % | BING VIDEOS | % |
|---|---|---|---|---|---|
| March | 18,557,201 | 5,717,222 | 30.81 | 3,922 | 0.02 |
| June | 10,274,903 | 930,941 | 9.06 | 2,147 | 0.02 |
| September | 6,683,502 | 842,801 | 12.61 | 1,567 | 0.02 |
| December | 5,436,832 | 310,845 | 5.72 | 1,139 | 0.02 |

Data shows clearly how Bing decreases its coverage along 2010 from 18,557,201 files in March to only 5,436,832 in September. The percentage of graphic files have been also reduced (from 30.81% to 5.72%), although the great value in March is due to the web behavior of "upc.es" (this issue is identified with the analysis per academic web domain, in the next chapter). Otherwise, the coverage of multimedia files is not representative in any shot along the period (0.02%).

**3.2. Count per academic web domain**

This chapter is divided into 3 sections, considering graphic, multimedia and blog content.

**3.2.1 Graphic count**

*a) Google images*

*Count and evolution*

Table 4 shows both raw data (total data retrieved from this search engine), normalized results obtained per sample, and Rc value (average of 4 normalized values) for all URLs of the Spanish university system (full details of the 141 URLs, including range and standard deviation, available in Annex I).

**Table 4. Ranking of URLs according to Rc(top 20)**

| WEB DOMAIN (n=141) | RAW | | | | NORMALIZED | | | | Rc |
|---|---|---|---|---|---|---|---|---|---|
| | MAR | JUN | SEP | DEC | MAR | JUN | SEP | DEC | |
| us.es | 90,500 | 96,100 | 134,000 | 147,000 | 4.81 | 4.99 | 6.19 | 6.38 | 5.59 |
| uv.es | 104,000 | 99,900 | 111,000 | 110,000 | 5.53 | 5.18 | 5.13 | 4.77 | 5.15 |
| ua.es | 89,000 | 94,100 | 103,000 | 109,000 | 4.73 | 4.88 | 4.76 | 4.73 | 4.78 |
| ucm.es | 84,800 | 82,300 | 74,800 | 80,900 | 4.51 | 4.27 | 3.46 | 3.51 | 3.94 |
| upc.edu | 51,200 | 64,200 | 79,700 | 94,300 | 2.72 | 3.33 | 3.68 | 4.09 | 3.46 |
| ugr.es | 66,700 | 65,100 | 73,400 | 73,400 | 3.55 | 3.38 | 3.39 | 3.18 | 3.37 |
| uab.es | 80,500 | 48,400 | 63,800 | 68,000 | 4.28 | 2.51 | 2.95 | 2.95 | 3.17 |
| ehu.es | 55,500 | 51,100 | 58,700 | 66,200 | 2.95 | 2.65 | 2.71 | 2.87 | 2.80 |
| upm.es | 53,400 | 53,100 | 54,800 | 64,300 | 2.84 | 2.76 | 2.53 | 2.79 | 2.73 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *upv.es* | 47,500 | 43,900 | 53,600 | 67,200 | 2.52 | 2.28 | 2.48 | 2.91 | **2.55** |
| *unizar.es* | 49,100 | 50,000 | 52,800 | 51,800 | 2.61 | 2.60 | 2.44 | 2.25 | **2.47** |
| *ub.es* | 37,300 | 50,000 | 56.200 | 52,200 | 1.98 | 2.60 | 2.60 | 2.26 | **2.36** |
| *um.es* | 43,200 | 41,800 | 49,800 | 49,500 | 2.30 | 2.17 | 2.30 | 2.15 | **2.23** |
| *ub.edu* | 37,400 | 38,500 | 49,500 | 50,400 | 1.99 | 2.00 | 2.29 | 2.19 | **2.11** |
| *uab.cat* | 554 | **61,300** | **57,700** | **59.400** | 0.03 | 3.18 | 2.67 | 2.58 | **2.11** |
| *unirioja.es* | 42,100 | 42,900 | 47,400 | 34,800 | 2.24 | 2.23 | 2.19 | 1.51 | **2.04** |
| *upc.es* | 37,200 | 40,000 | 40,600 | 46,400 | 1.98 | 2.08 | 1.88 | 2.01 | **1.99** |
| *uvigo.es* | 35,500 | 29,600 | 41,700 | 58,500 | 1.89 | 1.54 | 1.93 | 2.54 | **1.97** |
| *uam.es* | 37,100 | 38,800 | 40,700 | 45,200 | 1.97 | 2.01 | 1.88 | 1.96 | **1.96** |

The first three universities are the only which surpass the value of 100,000 files in the last sample, and generate jointly more than 15% of all content. On the other hand, it is surprising the low values retrieved from "unirioja.es", one of the largest Spanish academic web domains. Otherwise, some other unusual trends are detected. For example, the unexplained drop of "uab.es" from March to June data, which is complemented with an increment in "uab.cat" for the same period. This can be explained under political decisions to manage specific alternative TLDs.

The normalized data should not be interpreted in a longitudinal way (raw data should be used for that purpose), because it depends on the global count of the Spanish academic web space, and the interpretation of a concrete web domain is affected by the global performance of the web space. The aim of this normalization is just to show the proportion of number of pages of each web domain respect to the global count in each moment, and present later this value averaged at the end of the analyzed period.

To further analyze data variations over time, table 5 presents the URLs with major and minor range values over web domains along 2010.

**Table 5. URLs with major and minus range (R), and annual nominal interest rate (r)**

| WEB DOMAIN | Range (max) | r (%) | WEB DOMAIN | Range (min) | r (%) |
|---|---|---|---|---|---|
| *uab.cat* | 58,846 | 8.87 | upf.edu | -8,500 | -0.24 |
| *us.es* | 56,500 | 0.52 | udc.es | -9,100 | -0.29 |
| *upc.edu* | 43,100 | 0.66 | uah.es | -10,200 | -0.29 |
| *uvigo.es* | 23,000 | 0.53 | uab.es | -12,500 | -0.17 |
| *ua.es* | 20,000 | 0.21 | usc.es | -14,700 | -0.33 |

The range is generally positive for all web domains (only 24 domains record a negative range). The growth rate (r) also remarks the performance of "uab.cat", due to the low results in the first sample and the value obtained at the end (while R is based on subtracting, r is based on division). For the same reason, "upc.edu" achieves higher "r" value than "us.es"

Table 6 adds information about the 4 graphic formats studied (JPG, GIF, BMP, and PNG) for the top 10 URLs with more global graphic file count, indicating for each URL the Rc values obtained.

**Table 6. Rc for graphic files (JPG, GIF, BMP, PNG)**

| WEB DOMAIN | Rc (Global) | Rc (JPG) | Rc (GIF) | Rc (BMP) | Rc (PNG) |
|---|---|---|---|---|---|
| *us.es* | **5.59** | 5.67 | 5.22 | 2.22 | **13.25** |
| *uv.es* | **5.15** | 6.18 | 5.32 | **19.33** | 3.89 |
| *ua.es* | **4.78** | 5.42 | 4.94 | 2.14 | 4.17 |
| *ucm.es* | **3.94** | 3.58 | 4.64 | 2.41 | 8.20 |
| *upc.edu* | **3.46** | 2.93 | 3.42 | 1.55 | 3.42 |
| *ugr.es* | **3.37** | 3.19 | 4.18 | 1.72 | 2.64 |
| *uab.es* | **3.17** | 2.65 | 2.62 | 1.55 | 7.55 |
| *ehu.es* | **2.80** | 2.59 | 5.08 | 1.95 | 1.01 |
| *upm.es* | **2.73** | 2.53 | 3.41 | 1.97 | 4.70 |
| *upv.es* | **2.55** | 2.75 | 3.02 | 1.98 | 1.97 |

The usage drop of PNG files is mainly due to range values detected for the different domains belonging to UAB University: "uab.cat" (153,700) and "uab.es" (24,700). These figures indicate again a clear change in graphic files policy management at this university. Minimum and maximum ranges for each of the graphic formats are shown in table 7, for illustrative purposes.

**Table 7. Major and minor range (R) for graphic files count**

| | JPG | | |
|---|---|---|---|
| WEB DOMAIN | Range (max) | WEB DOMAIN | Range (min) |
| *us.es* | 35,100 | *uv.es* | -3,800 |
| *uab.cat* | 33,114 | *ull.es* | -7,180 |
| *upc.edu* | 17,800 | *udc.es* | -9,500 |
| *uvigo.es* | 11,600 | *uah.es* | -11,300 |
| *upv.es* | 10,500 | *urjc.es* | -49,560 |

| | GIF | | |
|---|---|---|---|
| WEB DOMAIN | Range (max) | WEB DOMAIN | Range (min) |
| *upc.edu* | 26,900 | *uib.es* | -2,190 |
| *uab.cat* | 15,300 | *uah.es* | -2,420 |
| *ua.es* | 13,900 | *ull.es* | -4,410 |
| *us.es* | 13,300 | *ucm.es* | -4,500 |
| *ub.es* | 10,700 | *ehu.es* | -9,700 |

| | BMP | | |
|---|---|---|---|
| WEB DOMAIN | Range (max) | WEB DOMAIN | Range (min) |
| *uv.es* | 1,047 | *ugr.es* | -19 |
| *uca.es* | 74 | *uch.ceu.es* | -31 |
| *uclm.com* | 65 | *ull.es* | -32 |
| *uclm.edu* | 60 | *usc.es* | -42 |
| *upf.edu* | 53 | *ulpgc.es* | -56 |

| | PNG | | |
|---|---|---|---|
| WEB DOMAIN | Range (max) | WEB DOMAIN | Range (min) |
| *upc.edu* | 13,979 | *universidadcamilojosecela.es* | -39 |
| *us.es* | 13,100 | *upf.es* | -94 |
| *ua.es* | 8,450 | *upf.edu* | -5,300 |
| *ub.edu* | 4,870 | *uab.es* | -24,700 |
| *ugr.es* | 4,290 | *uab.cat* | -153,700 |

*Inconsistencies*

Comparing the partial results for each of the types of graphic files and the global graphic count, a number of inconsistencies are observed. In order to analyze them in more detail we have proceeded to obtain for each URL the sum of the results from the 4 types of formats (TOT), and to compare them with the global results (TG).

Contrary to expectations, it is found a set of URLs in which the sum of the 4 graphic formats is superior respect to the global count, which shows some methodological weaknesses of this search functionality (previously detected in table 2 for accumulated count). The results obtained for the URLs where this phenomenon is found are offered in table 8 (data from December, 2010), where the errors obtained are particularly noted in the domains "uv.es", "ua.es", "ucm.es" and "us.es" (domains with large figures), where difference exceeds 10,000 results.

**Table 8. Inconsistencies between global graphic count (TG) and sumatory of graphic files (TOT)**

| WEB DOMAIN | TG | TOT | DIFFERENCE | WEB DOMAIN | TG | TOT | DIFFERENCE |
|---|---|---|---|---|---|---|---|
| universidadsanjorge.org | 7 | 8 | 1 | ugr.es | 73,400 | 76,041 | 2,641 |
| usj.es | 117 | 119 | 2 | uib.es | 24,000 | 27,171 | 3,171 |
| uoc.org | 86 | 91 | 5 | usc.es | 35,100 | 38,477 | 3,377 |
| upcomillas.net | 77 | 94 | 17 | unex.es | 21,200 | 24,861 | 3,661 |
| uniovi.es | 20,200 | 20,226 | 26 | usal.es | 44,100 | 48,092 | 3,992 |
| upcomillas.org | 63 | 98 | 35 | upc.es | 46,400 | 50,539 | 4,139 |
| url.cat | 190 | 243 | 53 | uam.es | 45,200 | 50,011 | 4,811 |
| uspceu.com | 785 | 841 | 56 | uva.es | 33,100 | 37,916 | 4,816 |
| urjc.net | 66 | 124 | 58 | upf.edu | 30,800 | 36,048 | 5,248 |
| uchceu.es | 2,090 | 2,159 | 69 | uab.es | 68,000 | 73,473 | 5,473 |
| unir.net | 74 | 152 | 78 | ehu.es | 66,200 | 71,684 | 5,484 |
| udg.edu | 28,500 | 28,580 | 80 | uned.es | 35,500 | 41,492 | 5,992 |
| upf.es | 1,400 | 1,669 | 269 | upv.es | 67,200 | 74,025 | 6,825 |
| udc.es | 26,000 | 26,428 | 428 | unizar.es | 51,800 | 58,865 | 7,065 |
| uoc.es | 3,150 | 3,624 | 474 | upc.edu | 94,300 | 101,484 | 7,184 |
| uco.es | 22,000 | 22,654 | 654 | uab.cat | 59,400 | 67,768 | 8,368 |
| ie.edu | 19,600 | 20,434 | 834 | ub.es | 52,200 | 60,870 | 8,670 |
| unican.es | 19,200 | 20,123 | 923 | um.es | 49,500 | 58,463 | 8,963 |
| uma.es | 30,000 | 30,936 | 936 | upm.es | 64,300 | 74,010 | 9,710 |
| uhu.es | 21,100 | 22,140 | 1,040 | ub.edu | 50,400 | 60,125 | 9,725 |
| ujaen.es | 13,700 | 14,905 | 1,205 | uv.es | 110,000 | 123,330 | 13,330 |
| unav.es | 36,900 | 38,723 | 1,823 | ua.es | 109,000 | 123,599 | 14,599 |
| uclm.es | 32,500 | 34,524 | 2,024 | ucm.es | 80,900 | 95,919 | 15,019 |
| ulpgc.es | 24,600 | 26,768 | 2,168 | us.es | 147,000 | 174,820 | 27,820 |

*Dispersion*

A dispersion of count between the different alias domains is also identified. Table 9 shows some of the most important cases detected.

**Table 9. Global graphic file count dispersion between alias web domains**

| WEB DOMAIN | GRAPHIC COUNT | WEB DOMAIN | GRAPHIC COUNT |
|---|---|---|---|
| nebrija.com | 1,140 | uniovi.com | 180 |
| nebrija.es | 827 | uniovi.es | 20,200 |
| uab.cat | 59,400 | uniovi.net | 1,100 |
| uab.es | 68,000 | uniovi.org | 0 |
| uao.cat | 807 | unioviedo.com | 0 |
| uao.es | 2,040 | unioviedo.es | 5,400 |
| ub.cat | 5,960 | unioviedo.net | 2 |
| ub.edu | 50,400 | unioviedo.org | 0 |
| ub.es | 52,200 | uoc.cat | 23 |
| uclm.com | 9,630 | uoc.edu | 36,400 |
| uclm.edu | 9,180 | uoc.es | 3,150 |
| uclm.es | 32,500 | uoc.org | 86 |
| uclm.net | 7,130 | upcomillas.com | 1,180 |
| uclm.org | 7,540 | upcomillas.edu | 19 |
| udg.cat | 71 | upcomillas.es | 6,530 |
| udg.edu | 28,500 | upcomillas.net | 77 |
| udg.es | 4,380 | upcomillas.org | 63 |
| ceuuch.es | 53 | urv.cat | 10,700 |
| uch.ceu.es | 9,420 | urv.es | 9,920 |
| uchceu.es | 2,090 | urv.net | 1,780 |

The region of Catalonia constitutes the more complex environment; universities like UAB, UB or URV maintain alias with similar and elevated results. It also highlights particularly the case of UCLM, where, in addition to the official domain "uclm.es", the rest of alias maintains similar and very high results sets.

b) *Bing images*

The problem of dispersion between URL alias is also detected in Bing; notwithstanding, in order not to be repetitive, data are omitted, and they can be consulted directly in the Annex II.

Table 10 shows top 20 URLs with higher Rc value (full details are also included in Annex II); it is identified a sharp drop in the number of results obtained mainly from the December sample, which confirms the loss of representation of this search engine, already observed in global count calculations.

**Table 10. Mean relative representation factor in count (Rc) (Bing images)**

| WEB DOMAIN (n=141) | RAW | | | | NORMALIZED | | | | Rc |
|---|---|---|---|---|---|---|---|---|---|
| | MAR | JUN | SEP | DEC | MAR | JUN | SEP | DEC | |
| upc.es | 4,460,000 (795.01%) | 29,500 (10.69%) | 20,100 (44.47%) | 6,400 (19.75%) | 78.01 | 3.17 | 2.38 | 2.06 | **21.41** |
| ua.es | 38,000 (9.38%) | 65,700 (**20.86%**) | 80,100 (**39.07%**) | 6,000 (4.14%) | 0.66 | 7.06 | 9.50 | 1.93 | **4.79** |
| uv.es | 73,400 (9.21%) | 68,800 (**22.78%**) | 65,600 (**19.24%**) | 6,190 (2.25%) | 1.28 | 7.39 | 7.78 | 1.99 | **4.61** |
| ucm.es | 34,300 (1.83%) | 41,600 (5.21%) | 40,900 (12.32%) | 5,410 (1.86%) | 0.60 | 4.47 | 4.85 | 1.74 | **2.92** |
| upv.es | 57,600 (8.66%) | 40,700 (12.60%) | 33,000 (**20.25%**) | 6,150 (5.08%) | 1.01 | 4.37 | 3.92 | 1.98 | **2.82** |
| ub.es | 31,200 (7.65%) | 40,600 (**42.83%**) | 32,600 (**34.39%**) | 5,860 (8.05%) | 0.55 | 4.36 | 3.87 | 1.89 | **2.67** |
| us.es | 72,400 (11.79%) | 28,800 (9.26%) | 36,300 (9.26%) | 6,040 (1.92%) | 1.27 | 3.09 | 4.31 | 1.94 | **2.65** |
| upm.es | 52,300 (4.84%) | 36,400 (7.79%) | 31,600 (16.37%) | 5,400 (3.33%) | 0.91 | 3.91 | 3.75 | 1.74 | **2.58** |
| ugr.es | 28,200 (5.72%) | 30,900 (9.12%) | 29,200 (12.32%) | 5,490 (2.72%) | 0.49 | 3.32 | 3.46 | 1.77 | **2.26** |
| uam.es | 31,700 (8.13%) | 30,900 (10.58%) | 24,900 (11.91%) | 5,700 (3.54%) | 0.55 | 3.32 | 2.95 | 1.83 | **2.17** |
| ehu.es | 33,300 (11.44%) | 29,600 (13.58%) | 25,800 (15.36%) | 5,530 (4.32%) | 0.58 | 3.18 | 3.06 | 1.78 | **2.15** |
| uah.es | 47,900 (17.94%) | 23,800 (17.00%) | 22,200 (**23.92%**) | 6,510 (9.80%) | 0.84 | 2.56 | 2.63 | 2.09 | **2.03** |
| unizar.es | 32,700 (3.40%) | 25,400 (5.06%) | 23,600 (11.57%) | 5,670 (3.61%) | 0.57 | 2.73 | 2.80 | 1.82 | **1.98** |
| udc.es | 27,300 (8.72%) | 22,800 (16.17%) | 19,000 (**32.37%**) | 6,670 (14.07%) | 0.48 | 2.45 | 2.25 | 2.15 | **1.83** |
| um.es | 31,200 (9.20%) | 22,100 (10.09%) | 19,200 (11.71%) | 5,750 (4.20%) | 0.55 | 2.37 | 2.28 | 1.85 | **1.76** |
| uclm.es | 21,900 (9.56%) | 21,400 (15.85%) | 19,100 (14.92%) | 6,280 (5.51%) | 0.38 | 2.30 | 2.27 | 2.02 | **1.74** |
| uab.es | 31,500 (4.97%) | 19,700 (12.63%) | 15,500 (14.09%) | 6,540 (7.86%) | 0.55 | 2.12 | 1.84 | 2.10 | **1.65** |
| uvigo.es | 19,400 (6.74%) | 19,300 (**20.44%**) | 16,100 (17.44%) | 6,350 (9.53%) | 0.34 | 2.07 | 1.91 | 2.04 | **1.59** |
| unex.es | 9,100 (10.10%) | 21,000 (**21.58%**) | 18,900 (**24.48%**) | 5,030 (16.49%) | 0.16 | 2.26 | 2.24 | 1.62 | **1.57** |
| uji.es | 29,400 (8.80%) | 18,200 (13.28%) | 14,600 (**20.00%**) | 6,100 (9.37%) | 0.51 | 1.96 | 1.73 | 1.96 | **1.54** |

The presence of domains belonging to polytechnic universities should be pointed out, jointly with the low performance (as detected with Google) for "unirioja.es", which show 5,710 graphic files in December (1.34% of the global count for this domain: 427,000), and hence does not appear in table 10.

Special attention must also be paid to the high value obtained by "upc.es" in March (and its gradual decline), which clearly shows an inconsistency. This anomalous result in March (which causes a partial Rc= 78.01), and the very low value recovered in the following shot (Rc= 3.17), provokes an increase of ratios in the rest of domains. View this in a longitudinal way has no sense and shows that all the domains except "upc.es" increase their values when in fact this is not true. For this reason is recommended only to take into account the last 3 shots, and the raw values.

For each raw data is also provided their percentage respect to the total number of pages for this web domain, in order to give an overview about the proportion of graphic files in each academic web domain. Anyway, the sharp decrease of coverage generates several unexpected results, especially in the June shot.

In fact, the global graphic file (accumulated) count throughout the whole Spanish academic web space goes from 5,717,222 results in March to 310,845 on December (negative statistic range of 5,406,377). Figure 8 illustrates this negative trend, comparing the global count distribution of the last two samples (September and December).
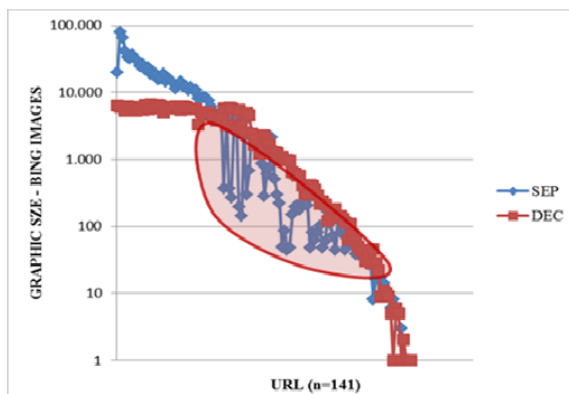


**Figure 8. Distribution of global graphic file accumulated count on September and December, 2010 (*Bing images*)**

On one hand, the largest URLs appear (in December sample) in a very narrow range of values, which is reflected in the almost zero slope of the distribution on the left side of the table. On the other hand, the URLs which are basically losing weight are those with higher global graphic file count (with certain exceptions not justified, such as "uco.es", "unav.es", "unirioja.es" or "uhu.es").

However, there are a wide range of domains increasing their results from September to December. In all of these URLs, a sharp drop is detected in June, rising again in December (although without reaching the initial values, i.e., presenting a negative statistical range), just when the most important domains fall.

Besides these web domains, it is detected another group of URLs, with discrete graphic counts, which have a positive range. All of them are shown in table 11.

**Table 11. URLs with positive range (Bing images)**

| WEB DOMAIN | MAR | JUN | SEP | DIC | Range | r (%) |
|---|---|---|---|---|---|---|
| upsa.es | 595 | 188 | 151 | 644 | 49 | 0.08 |
| uvic.cat | 90 | 114 | 125 | 139 | 49 | 0.46 |
| uao.es | 341 | 211 | 208 | 404 | 63 | 0.17 |
| nebrija.com | 842 | 79 | 48 | 941 | 99 | 0.11 |
| ufv.es | 937 | 83 | 49 | 1.110 | 173 | -3.26 |
| uemc.es | 225 | 79 | 81 | 406 | 181 | 0.64 |
| upco.es | 945 | 408 | 2,130 | 1,310 | 365 | -3.23 |
| udl.cat | 1,290 | 896 | 743 | 1,670 | 380 | 0.27 |
| urv.cat | 823 | 658 | 511 | 1,280 | 457 | -3.21 |
| uniovedo.es | 169 | 78 | 44 | 983 | 814 | 2.21 |

**c) Comparative**

Figure 9 provides a comparative between the global results in *Google images*, *Bing images*, and *Bing*.
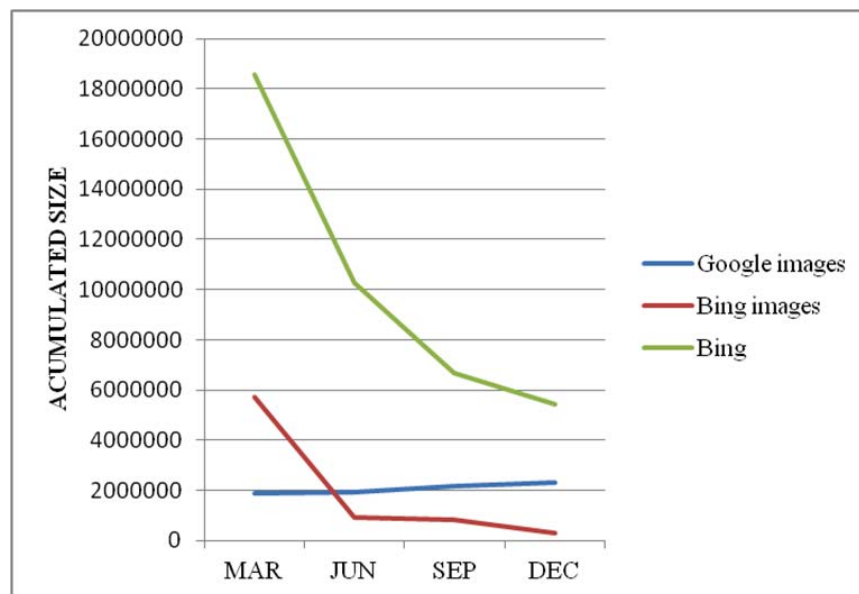


**Figure 9. Comparison between the distribution of global graphic count and global count both for *Bing* and *Google* (December, 2010)**

A sharp coverage drop in *Bing* (reflected on *Bing images*) is detected, while the results of *Google images* remain more or less constant for the taken period.

Despite the inconsistencies between *Bing images* results throughout the measurements period, if we take as reference the most recent data (December 2010), the similarities between *Google images* and *Bing images* are higher than expected.

Figure 10 shows the comparative distribution of both sets of data, where it is identified a positive correlation between both sources, except for two important areas. First one is detected in the middle "x" axis (where low results imply small overall differences), and the other one is detected on the upper zone, where less coverage is detected on *Bing images*. For example, "us.es" gets 147,000 hits on *Google images*, for just 6,040 images in *Bing images*. Other URLs with large differences include, among others, "uv.es" (110,000 and 6,190 respectively) and "ua.es" (100,000 and 6,000).
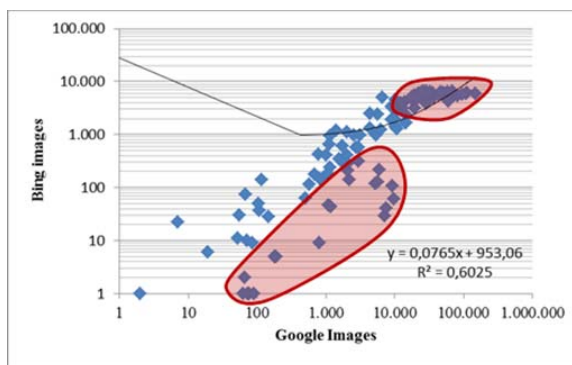


**Figure 10.** Correlation between *Bing images* y *Google images* (December 2010)

On the other hand, the low performance of *Bing images* in the upper zone seems to show a limitation in the image retrieving process (does not display more than 10,000 results per site). This phenomenon is only detected in December shot.

**Global inconsistencies**

A comparison between the global search engine (*Bing*) and the specific image search engines (*Bing images* and *Google images*) allow identifying certain inconsistences results. Below (table 12) is presented a comparison between the results of *Bing* and *Google images*, captured in the last sample (December 2010). The coverage difference between the two sources allows -for some URLs- getting more search results in the specific image query than in the global query.

**Table 12**. Comparison between global count (*Bing*) y global graphic count (*Google images*) (Dec 2010)

| WEB DOMAIN | GOOGLE IMAGES | BING | WEB DOMAIN | GOOGLE IMAGES | BING |
|---|---|---|---|---|---|
| upc.cat | 2 | 0 | url.es | 500 | 366 |
| unirioja.org | 3 | 1 | uniovi.com | 180 | 6 |
| uoc.cat | 23 | 16 | uoc.es | 3,150 | 2,940 |
| uic.cat | 8 | 1 | ucv.es | 2,060 | 1,490 |
| upcomillas.edu | 19 | 11 | urv.net | 1,780 | 1,170 |
| upf.cat | **16** | 0 | uao.cat | 807 | 14 |
| uimp.net | 20 | 0 | upcomillas.com | 1,180 | 116 |
| unica.es | 20 | 0 | udl.cat | 14,300 | 13,100 |
| ceuuch.es | 53 | 17 | uchceu.es | 2,090 | 674 |
| universidadcamilojosecela.es | 55 | 16 | upco.es | 4,370 | 2,690 |
| uemc.edu | 681 | 639 | uib.cat | 5,220 | 2,000 |
| uimp.org | 45 | 0 | unav.edu | 5,660 | 372 |
| udg.cat | 71 | 19 | ub.cat | 5,960 | 500 |
| upcomillas.org | 63 | 8 | uclm.net | **7,130** | 73 |
| urjc.net | 66 | 2 | uclm.org | **7,540** | 59 |
| upcomillas.net | 77 | 7 | uclm.edu | **9,180** | 132 |
| uoc.org | 86 | 10 | uclm.com | **9,630** | 308 |
| fundacionviu.es | 89 | 6 | uab.cat | **59,400** | 47,800 |
| unica.edu | 105 | 18 | upc.es | **46,400** | 32,400 |
| udl.es | 17,700 | 17,600 | upc.edu | **94,300** | 79,700 |
| url.cat | 190 | 66 | | | |

Comparing the counts obtained both from *Bing images* and *Bing*, this phenomenon practically disappears, with only 2 URLs detected with this problem: "universidadcamilojosecela.es" (16 results for *Bing*, and 30 for *Bing images*), and "unica.edu" (18 results for *Bing*, and 50 in *Bing images*).

In fact, the correlation between *Bing* and *Bing images* is quite large, as reflected in figure 11, which only differs fundamentally on the area of high performance in both search engines (URLs with larger global *Bing* count also are the URLs with more global graphic count).
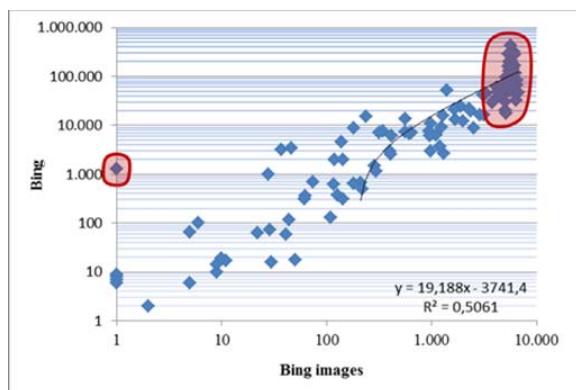


**Figure 11.** Correlation between *Bing* and *Bing images* (Dec 2010)

**3.2.2. Multimedia count**

Tables with complete data collected both for *Google videos* and *Bing videos* are available in the corresponding annexes III and IV.

a) *Google videos*

The top 20 academic web domains with higher Rc value for the Global multimedia file count indicator are shown in table 13.

**Table 13. Mean relative representation factor in count (Rc) (*Google videos*)**

| WEB DOMAIN (n=141) | RAW | | | | NORMALIZED | | | | Rc |
|---|---|---|---|---|---|---|---|---|---|
| | MAR | JUN | SEP | DEC | MAR | JUN | SEP | DEC | |
| upc.edu | 649 | 581 | 536 | 646 | 40.11 | 37.22 | 49.31 | 37.23 | **40.97** |
| uva.es | 138 | 64 | 56 | 216 | 8.53 | 4.10 | 5.15 | 12.45 | **7.56** |
| us.es | 116 | 113 | 79 | 49 | 7.17 | 7.24 | 7.27 | 2.82 | **6.13** |
| unav.es | 65 | 64 | 64 | 83 | 4.02 | 4.10 | 5.89 | 4.78 | **4.70** |
| upv.es | 46 | 48 | 39 | 125 | 2.84 | 3.07 | 3.59 | 7.20 | **4.18** |
| uc3m.es | 37 | 26 | 71 | 53 | 2.29 | 1.67 | 6.53 | 3.05 | **3.38** |
| uvigo.es | 14 | 20 | 9 | 146 | 0.87 | 1.28 | 0.83 | 8.41 | **2.85** |
| udg.edu | 15 | 132 | 5 | 6 | 0.93 | 8.46 | 0.46 | 0.35 | **2.55** |
| upm.es | 68 | 33 | 7 | 33 | 4.20 | 2.11 | 0.64 | 1.90 | **2.22** |
| uniovi.es | 55 | 54 | 5 | 10 | 3.40 | 3.46 | 0.46 | 0.58 | **1.97** |
| ub.edu | 25 | 30 | 26 | 33 | 1.55 | 1.92 | 2.39 | 1.90 | **1.94** |
| uclm.es | 39 | 40 | 15 | 16 | 2.41 | 2.56 | 1.38 | 0.92 | **1.82** |
| ehu.es | 21 | 29 | 14 | 13 | 1.30 | 1.86 | 1.29 | 0.75 | **1.30** |
| uned.es | 15 | 17 | 12 | 33 | 0.93 | 1.09 | 1.10 | 1.90 | **1.26** |
| ufv.es | 21 | 23 | 12 | 17 | 1.30 | 1.47 | 1.10 | 0.98 | **1.21** |
| um.es | 17 | 18 | 12 | 5 | 1.05 | 1.15 | 1.10 | 0.29 | **0.90** |
| ucm.es | 15 | 20 | 7 | 7 | 0.93 | 1.28 | 0.64 | 0.40 | **0.81** |
| uca.es | 9 | 13 | 7 | 19 | 0.56 | 0.83 | 0.64 | 1.10 | **0.78** |
| ugr.es | 19 | 10 | 3 | 17 | 1.17 | 0.64 | 0.28 | 0.98 | **0.77** |
| unizar.es | 12 | 15 | 5 | 15 | 0.74 | 0.96 | 0.46 | 0.86 | **0.76** |

The results show a high representation of the domain "upc.edu" in all samples. Other domains such as "uva.es" and "upv.es" appear at the top due to the significant increase recorded in December. In any case, the values are discrete. In the last sample, a total of 83 URLs do not obtain any result, while there are only 31 URLs with more than 5 results.

The global data also shows irregularities in their evolution over time. Specifically, it is detected a significant drop in June (where the accumulated multimedia count amounted to 1,561 results) to September (where only 1,087 results were obtained). This drop is mainly produced in "us.es" (which goes from 113 documents to 79) and "udg.edu" (from 132 to 5).

In December results grow again, although this is not due to the recovery of the domains described above but to the significant increase recorded in "upc.edu" (from 536 to 646), "uva.es" (56 to 216), "upv.es" (39 to 125), and "uvigo.es" (9 to 146).

b) *Bing videos*

As for table 13, table 14 shows the corresponding top 20 web domains that achieve greater Rc value on Bing videos.

**Table 14. Mean relative representation factor in count (Rc) (Bing videos)**

| WEB DOMAIN (n=141) | RAW | | | | NORMALIZED | | | | Rc |
|---|---|---|---|---|---|---|---|---|---|
| | MAR | JUN | SEP | DEC | MAR | JUN | SEP | DEC | |
| upc.edu | 988 (0.23%) | 568 (0.49%) | 495 (0.53%) | 414 (0.52%) | 25.19 | 26.46 | 31.59 | 36.35 | **29.90** |
| upv.es | 395 (0.06%) | 212 (0.07%) | 165 (0.10%) | 122 (0.10%) | 10.07 | 9.87 | 10.53 | 10.71 | **10.30** |
| **uab.es** | 141 (0.02%) | 231 (0.15%) | 62 (0.06%) | 12 (0.01%) | 3.60 | 10.76 | 3.96 | 1.05 | **4.84** |
| uclm.es | 151 (0.07%) | 106 (0.08%) | 68 (0.05%) | 59 (0.05%) | 3.85 | 4.94 | 4.34 | 5.18 | **4.58** |
| uv.es | 124 (0.02%) | 78 (0.03%) | 74 (0.02%) | 70 (0.03%) | 3.16 | 3.63 | 4.72 | 6.15 | **4.42** |
| unia.es | 94 (0.27%) | 79 (0.50%) | 50 (0.53%) | 33 (0.48%) | 2.40 | 3.68 | 3.19 | 2.90 | **3.04** |
| uc3m.es | 129 (0.03%) | 48 (0.04%) | 42 (0.04%) | 32 (0.03%) | 3.29 | 2.24 | 2.68 | 2.81 | **2.75** |
| unizar.es | 52 (0.01%) | 47 (0.01%) | 43 (0.02%) | 39 (0.02%) | 1.33 | 2.19 | 2.74 | 3.42 | **2.42** |
| **upm.es** | 127 (0.01%) | 38 (0.01%) | 35 (0.02%) | 21 (0.01%) | 3.24 | 1.77 | 2.23 | 1.84 | **2.27** |
| uam.es | 72 (0.02%) | 39 (0.01%) | 39 (0.02%) | 28 (0.02%) | 1.84 | 1.82 | 2.49 | 2.46 | **2.15** |
| udc.es | 41 (0.01%) | 43 (0.03%) | 37 (0.06%) | 35 (0.07%) | 1.05 | 2.00 | 2.36 | 3.07 | **2.12** |
| udg.edu | 31 (0.02%) | 33 (0.05%) | 26 (0.07%) | 18 (0.06%) | 0.79 | 1.54 | 1.66 | 1.58 | **1.39** |
| upc.es | 64 (0.01%) | 32 (0.01%) | 32 (0.07%) | 3 (0.01%) | 1.63 | 1.49 | 2.04 | 0.26 | **1.36** |
| uma.es | 35 (0.01%) | 24 (0.02%) | 24 (0.02%) | 21 (0.02%) | 0.89 | 1.12 | 1.53 | 1.84 | **1.35** |
| uoc.edu | 7 (0.00%) | 98 (0.07%) | 1 (0.00%) | 1 (0.00%) | 0.18 | 4.56 | 0.06 | 0.09 | **1.22** |
| uib.es | 47 (0.01%) | 26 (0.02%) | 15 (0.02%) | 16 (0.03%) | 1.20 | 1.21 | 0.96 | 1.40 | **1.19** |
| ehu.es | 71 (0.02%) | 21 (0.01%) | 17 (0.01%) | 8 (0.01%) | 1.81 | 0.98 | 1.08 | 0.70 | **1.14** |
| **uji.es** | 111 (0.03%) | 13 (0.01%) | 11 (0.02%) | 4 (0.01%) | 2.83 | 0.61 | 0.70 | 0.35 | **1.12** |
| umh.es | 42 (0.01%) | 17 (0.02%) | 18 (0.03%) | 16 (0.04%) | 1.07 | 0.79 | 1.15 | 1.40 | **1.10** |
| upf.edu | 65 (0.01%) | 28 (0.02%) | 17 (0.02%) | 4 (0.01%) | 1.66 | 1.30 | 1.08 | 0.35 | **1.10** |

Like Google videos, the largest domain is "upc.edu", although in this case the relative count obtained is lower (29.90, compared to 40.97 achieved on Google videos). Despite this agreement, the difference between these two sources is important because of the low values generally obtained by the different domains, which causes great changes in the positions of URL in each Rc ranking.

c) *Comparative*

There are domains with higher representation on Bing than Google, like "unavarra.es" or "upco.es", and domains far better positioned in Google. The most extreme case is "uva.es", which gets 2nd position on Google videos Rc ranking, while only does 51st on Bing videos.

Apart from the differences between sources, the volume of general results obtained by Bing videos is very discrete. In December 2010, a total of 80 URLs do not show any results. In addition, there is an excessive negative statistical range in almost all domains.

In fact, only 2 URLs are identified with a positive range ("upco.es", and "udl.cat"), although in small amounts. This fall of results can be visualized in figure 10, which compares the evolution of accumulated global multimedia count obtained for all the URLs in the two sources analyzed. In just nine months Bing videos has gone from almost 4,000 to recover just over 1,000 in December, where harvest less files than Google videos for the first time in the period.
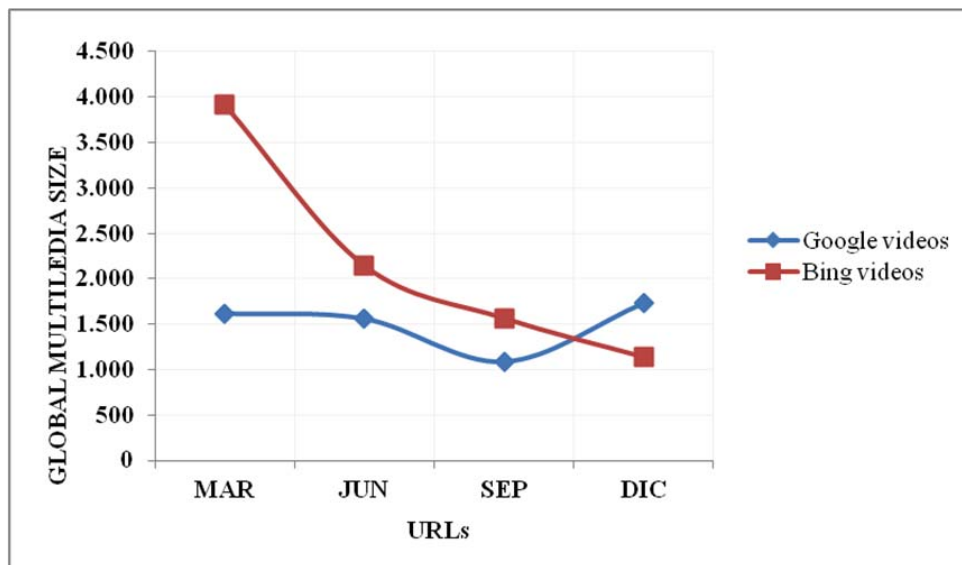
**Figure 12. Accumulated global multimedia file count per source and sample**

**3.2.3. Blog count (Google blogs)**

The blog content web space is formed (as of December 2010) by 337,845 results, which is an increase of 42,449 records from March 2010, when the first data collection was retrieved. Nevertheless, 59 URLs do not have any results, reflecting a highly skewed distribution.

Table 15 details the web domains with a value greater than 1 Rc in the period of study, as well as raw and normalized values for these URLs. A completed table with full details is available in Annex V.

**Table 15. Mean relative representation factor (Rc) (Google blogs)**

| WEB DOMAIN (n=141) | RAW | | | | NORMALIZED | | | | Rc |
|---|---|---|---|---|---|---|---|---|---|
| | MAR | JUN | SEP | DEC | MAR | JUN | SEP | DEC | |
| us.es | 25,569 | **29,754** | 32,634 | 32,340 | 9 | 10 | 10 | 10 | 9 |
| upf.edu | 21,385 | 21,194 | 21,105 | 19,085 | 7 | 7 | 6 | 6 | 6 |
| usc.es | 19,474 | 19,568 | 20,071 | 19,755 | 7 | 6 | 6 | 6 | 6 |
| usal.es | 15,855 | 16,417 | 18,005 | 19,524 | 5 | 5 | 5 | 6 | 5 |
| uv.es | 15,460 | 16,063 | 16,570 | 18,704 | 5 | 5 | 5 | 6 | 5 |
| ua.es | 16,413 | 16,381 | 16,252 | 17,488 | 6 | 5 | 5 | 5 | 5 |
| uva.es | 16,486 | 16,793 | 16,167 | 15,948 | 6 | 5 | 5 | 5 | 5 |
| ulpgc.es | 13,427 | 14,340 | 15,129 | 15,908 | 5 | 5 | 5 | 5 | 5 |
| udc.es | 13,959 | 14,623 | 14,516 | 14,249 | 5 | 5 | 4 | 4 | 4 |
| upm.es | 13,919 | 14,115 | 14,358 | 14,753 | 5 | 5 | 4 | 4 | 4 |
| upc.es | 12,832 | 13,792 | 14,584 | 13,766 | 4 | 4 | 4 | 4 | 4 |
| um.es | 12,095 | 13,313 | 13,877 | 15,060 | 4 | 4 | 4 | 4 | 4 |
| ie.edu | 11,530 | 11,330 | 12,184 | 12,698 | 4 | 4 | 4 | 4 | 4 |
| upv.es | 7,322 | 7,391 | 7,818 | 8,805 | 2 | 2 | 2 | 3 | 2 |
| uoc.edu | 6,683 | 7,114 | 7,439 | 7,770 | 2 | 2 | 2 | 2 | 2 |
| uco.es | 6,032 | 6,252 | 6,130 | **4,877** | 2 | 2 | 2 | 1 | 2 |
| ugr.es | 5,276 | 5,762 | 5,914 | 6,238 | 2 | 2 | 2 | 2 | 2 |
| umh.es | 5,050 | 4,719 | 5,129 | 6,909 | 2 | 2 | 2 | 2 | 2 |
| ucm.es | 4,984 | 5,097 | 5,101 | 4,678 | 2 | 2 | 2 | 1 | 2 |

The first position is occupied prominently by "us.es", with a relatively high representation value (9), indicating a highly distributed distribution. The UCM and UR universities (the other 2 major domains in global count) perform far behind, especially "unirioja.es" with only 32 results.

Otherwise, should be mentioned the presence of 2 private universities (UOC and IE) in the top places, which confirms the better performance of this type of content in private institutions.

The results, taking into account its upward trend, present strange behaviors in some URLs:

- The growth from March to June of "us.es" (from 25,569 to 29,754) and "ull.es" (from 1,331 to 3,813), and from June to September of "uma.es" (from 2,663 to 5,583) and "uah.es" (from just a result to 8,818). Also "deusto.es" should be remarked, with a growth from 1,846 to 4,022 from September to December.

- On the other, there are identified strong drops in December of "uco.es" (from 6,130 to 4,877), and "uab.cat" (from 5,026 to 551).

Despite these specific dysfunctions, and taking into account the upward trend in the data, the values show a high correlation between samples, as reflected in the distribution of results shown in figure 13.
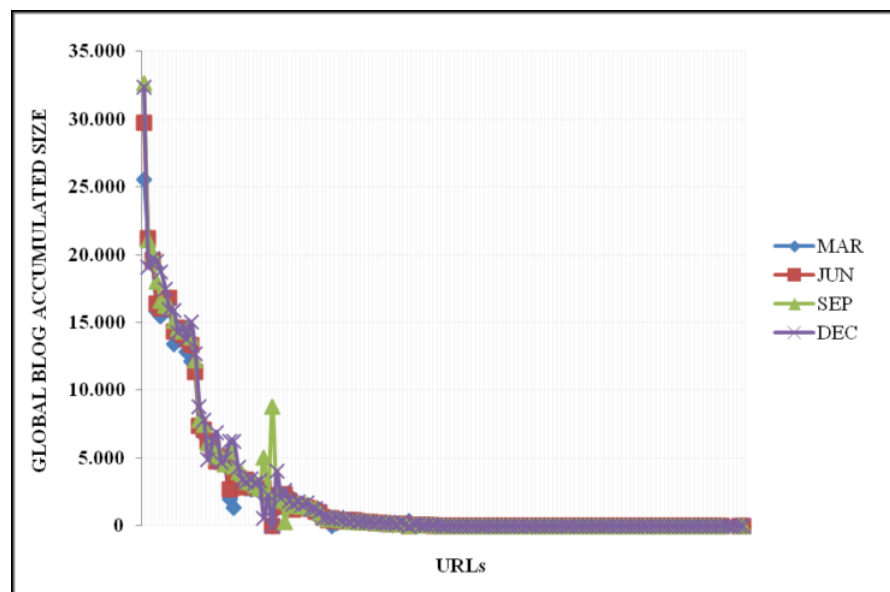
**Figure 13. Global blog accumulated count distribution per sample (Google videos)**

The amount of blog-type contents contrast with the few blog platforms identified within academic websites in the Spanish system. In December 2010 only 29 universities (of 76) hosted official academic blog platforms (table 16):

**Table 16. Blog platforms within Spanish academic websites (December 2010)**

| UNIVERSITY | BLOG PLATFORMS | WEB DOMAIN |
|---|---|---|
| USP | Banco de Talento | uspceu.com/blogs<br>uspceu.es/blogs |
| UAL | BLOG UAL | blog.ual.es<br>blog.ual.es:444 |
| UEM | Blog UEM | comunidad.uem.es/blog<br>comunidad.uem.es/blogs |
| US | Blog.us.es | blog.us.es |
| UOC | Blogs (UOC) | blogs.uoc.edu<br>uoc.edu/portal/castellano/difusio_i_publicacions/blogs<br>uoc.edu/portal/catala/difusio_i_publicacions/blogs<br>uoc.es/portal/castellano/difusio_i_publicacions/blogs<br>uoc.es/portal/catala/difusio_i_publicacions/blogs |
| UPC | Blogs (UPC) | blog.upc.edu |
| UDL | Blogs a la UDL | blogs.udl.cat |
| UCH | Blogs CEU | uch.ceu.es/principal/BlogsCEU<br>uchceu.es/principal/BlogsCEU |
| UAB | Blogs de la UAB | blogs.uab.cat<br>blogs.uab.es |
| UNAV | Blogs de la Universidad de Navarra | unav.edu/blogs<br>unav.es/blogs |
| UVA | Blogs de la Universidad de Valladolid | blogs.uva.es |
| UAN | Blogs de la Universidad Nebrija | blogs.nebrija.es |
| UV | Blogs de la Universitat de València | blogs.uv.es |
| UA | Blogs UA | blogs.ua.es |
| UAX | Blogs UAX | uax.es/blogs |
| UIMP | BLOGS UIMP | uimp.es/blogs |
| UMH | Blogs.umh.es | blogs.umh.es |
| UCJC | Comunidad de Blogs de la UCJC | ucjc.edu/blogs<br>ucjc.es/blogs<br>universidadcamilojosecela.es/blogs |
| UDE | DeustoBlog | blogs.deusto.es |
| USAL | Diarium: gestor de blogs | diarium.usal.es |
| EHU | EHUsfera | ehu.es/ehusfera |
| IE | IE Blogs | blogs.ie.edu |
| MU | Mondragon Unibertsitateko Blogak | blogs.mondragon.edu |
| UPV | Poliblogs | blogs.upv.es |
| UJA | Servicio de Blogs | blogs.ujaen.es |
| UNIA | Sistema de Blogs de la UNIA | blogs.unia.es |
| UDG | UdG Blogs | udg.edu/udgblogs<br>udg.es/udgblogs |
| ULL | UDV Blogs | blogs.udv.ull.es |
| UVIC | Uvic Blog | blocs.uvic.cat |

### 4. Conclusions

The main results obtained are showed below, structured by type of count analyzed: graphic, multimedia, and blog-type content.

*Graphic count*

The results obtained have allowed knowing the volume of indexed images in each domain as well as the major formats used, and their visibility in search engines, providing useful information on the management of image files in universities.

As regards the global graphic file count, "us.es" and "uv.es" URLs are the most representative domains. The absence of "unirioja.es" from the top should be pointed out, which shows the preponderance of other files on their servers (rich files, essentially).

If we analyze the files formats, JPG is the most used in university platforms, followed by the GIF and PNG files, whereas the BMP format is the least used. Previous results obtained by Baeza-Yates & Graells, (2007) and Tolosa et al. (2007), showed GIF as the most linked graphic format in Argentina and Chile web-spaces, of what can be inferred a widely use (count) of this format.

Results obtained by Alonso, Figuerola and Zazo (2004) also pointed to GIF as the most widely used, while the PNG format was not

analyzed. Although the coverage of different URLs and different method of analysis (that study used a self-spider instead of a commercial search engine) limit the possible comparisons, a decrease of GIF and an expansion of JPEG are observed.

With respect the coverage of the various search engines used, *Google images* gives higher results than *Bing images*. Moreover, the results obtained correlate well from both sources for URLs with larger amount of graphic files, while they differ in those with few files.

In any case, serious inconsistencies are detected in the behavior of search engines, especially *Google images*. In particular the results are higher in specific formats queries than in global graphic count.

The University of Valencia constitutes an example, for which 110,000 images are retrieved within the domain "uv.es" (December 2010, global graphic query), although the summary of images in specific formats (JPG, GIF, BMP and PNG) amounts to 123,330 images.

Additionally, inconsistencies between global search engines and image searches are detected: 41 URLs on *Google images* provides more results than *Bing* (global search).

Otherwise, a possible limitation in the image retrieving process in *Bing images* (estimated at 10,000 files for December 2010) could affect the performances of higher web sites, and correlation values among search engines. In any case, at present (2012), this limitation has disappeared.

These facts put into serious question the accuracy of global graphic count queries in search engines so that should not be recommended to use with metric purposes without caution.

*Multimedia count*

Although the number of universities with media platforms is high, the vision of them through search engines is not optimal. The number of media files is very small both on *Google videos* and *Bing videos*. In addition, universities with the largest number of indexed results in the first samples have reduced the number of hits over the period.

In any case, these results are logical in the case of *Google videos*. The service, despite stop working since April 2011 (after measurements made in this paper), stopped accepting new video files in May 2009. In the case of *Bing*, the loss of coverage joins the already detected in the global and format file count indicators.

In any case, the low performance of multimedia files should also be due to the growing trend to upload these files into platforms such as *Youtube*, outside the limits of academic websites, and the inaccessibility to some multimedia teaching material, deposited in academic intranets, and hence not harvested by commercial search engines.

*Blog count*

The results show, in general terms, an upward trend, indicating an increased use of blogs within the university platforms. Moreover, the results between samples show a high correlation and a reduced variability. Particularly noteworthy are the behavior of the following domains: "us.es" (which also over performs in graphic count), "upf.edu", and "usc.es".

Additionally, the following considerations should be remarked:

- Google Blogs recover large amounts of results at universities that do not have blogging platforms, which mean that universities generate such content without expressly provide a social platform for blogging.

- On the other hand, highlights the presence of public universities on the top, where the percentage of private universities with blogging platforms (and thus with a clear policy about this type of content) is significant. The large number of pages in public universities may explain this behavior.

As a final point, we conclude that graphic, multimedia and blog-type content files, counted together, represent a significant proportion of the Spanish academic websites, and so they must be taken into account in calculations of the total number of pages, but low accuracy of searchers (fundamentally image, and video) must be taken into account.

Respect to the first conclusion (proportion within Spanish academic web system):

Results show that for all three types of content analyzed, three different university groups are achieving high performances: polytechnic universities (UPC, UPM, and UPV), old established, multidisciplinary, and big universities (such as UCM, UB and UV), and small and specialized universities (such as UA, UPF and UVIGO). This confirms that not only the size of the university is influencing in their graphic, multimedia, and blog performance on the Web.

Multimedia and blog content can be indirect indicators of some academic activities (such as research, teaching and transfer supporting), whereas graphic files can reflect a compromise publishing digital collections or representing complementary material of web publications, for example.

The significant percentage of retrieved contents (especially images and blogs) indicates the need to detect their source, motivation and distribution within the academic web-space. Future work is desirable in order to identify the source of this content (digital collections, learning objects, blogs supporting lectures, news, research or other activities), to estimate their influence in the global academic web performance, and to calculate their possible correlation with other web indicators.

Respect to the second conclusion (low accuracy of image search engines):

Both *Google* and *Bing* present severe inconsistencies and irregularities in their image and video searches, which limit the capabilities of these tools to quantitative uses.

*Google* provides inconsistencies among graphic files count and global graphic count. Moreover, the inability of calculating global count accurately makes no possible to calculate the proportion of graphic, multimedia and blogs respect to the site size. On the other hand, *Google*'s coverage is bigger than *Bing*, and their evolution over time is also more coherent.

The period analyzed corresponds with the starting time of fusion with *Yahoo! Search*. This fact can explain the great instability and low coverage found. For that reason, all data recovered from *Bing* is showing indirectly the effects of fusion of both search engines in the Spanish academic web space. In any case, the strength of *Bing* is the possibility of calculating the proportion of graphic count respect to site size, but the disadvantages are the impossibility of searching specific graphic files, and the lower coverage respect to *Google*.

## 5. References

Aguillo, Isidro F. (2009). Measuring the institutions' footprint in the web. **Library Hi Tech**, 27(4): 540-556.

Aguillo, Isidro F.; Granadino, B.; Ortega, José L. & Prieto, J.A. (2006). Scientific research activity and communication measured with cybermetrics indicators. **Journal of the American Society for information science and technology**, 57(10): 1296-1302.

Aguillo, Isidro F.; Ortega, J.L. & Fernandez, M. (2008). Webometric Ranking of World Universities: Introduction, methodology, and future developments. **Higher Education in Europe**, 33 (2/3):233-244.

Alonso Berrocal, J.L.; Figuerola, C.G. & Zazo, A.F. (2004). Cibermetría: nuevas técnicas de estudio aplicables al Web. Trea, Gijón.

Angus, E. & Thelwall, M. (2010). Motivations for image publishing and tagging on Flickr. In: Turid Hedlund and Yasar Tonta (Eds.), Proceedings of the 14th International Conference on Electronic Publishing, Hanken School of Economics, Helsinki, 189-204).

Angus, E., Thelwall, M. & Stuart, D. (2008). General patterns of tag usage among university groups in Flickr. **Online Information**

Review, 32 (1): 89-101.

Angus, E.;Thelwall, M. & Stuart, D. (2010). Flickr's potential as an academic image resource: an exploratory study. **Journal of Librarianship and Information Science**, 42(4): 268–278.

Baeza-Yates, R.; & Graells, E. (2007). Características de la web chilena 2007. <http://www.ciw.cl/caracterizacion-web/estudio2007>

Cha, M.; Kwak, H.; Rodriguez, P.; Ahn, Y-Y. & Moon, S. (2009). Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems. **IEEE/ACM Transactions on Networking**, 17 (5): 1357-1370.

Chen, H. (2001). An analysis of image queries in the field of art history. **Journal of the American Society for Information Science and Technology**, 52(3): 260–273.

Choi , Y. & Rasmussen , E. (2003). Searching for images: The analysis of users' queries for image retrieval in American history. **Journal of the American Society for Information Science and Technology**, 54(6): 498–511.

Goodfellow, T. & Graham, S. (2007). The Blog as a High-Impact Institutional Communication Tool. **Electronic Library**, 25(4): 395-400.

Goodrum, A. & Spink, A. (2001). Image searching on the Excite search engine. **Information Processing & Management**, 37(2): 295–312.

Jansen, B. (2008). Searching for digital images on the web. **Journal of Documentation**, 64(1): 81–101.

Koehler, W. (1999). An analysis of web page and web site constancy and permanence. **Journal of the American Society for Information Science**, 50(2): 162–180.

Kousha, K. & Thelwall. M. (2008). Assessing the impact of research on teaching: an automatic analysis of online syllabuses in science and social sciences. **Journal of the American Society of Information Science and Technology**, 59(13): 2060-2069.

Kousha, K.; Thelwall, M.; Abdoli, M. (in press). The role of online videos in research communication: a content analysis of YouTube videos cited in academic publications. **Journal of the American Society for Information Science and Technology**.

Kousha, K.; Thelwall, M. & Rezaie, S. (in press). Can the impact of scholarly images be assessed online? An exploratory study using image identification technology. **Journal of the American Society for Information Science and Technology**.

Kousha, K.; Thelwall, M.; Rezaie, E. (2010). Using the Web for Research evaluation: The integrated online impact indicator. **Journal of informetrics**, 4(1): 124-135.

Orduña-Malea, E.; Serrano-Cobos, J.; Lloret-Romero, N. (2009). Las universidades públicas españolas en Google Scholar: presencia y evolución de su publicación académica web. **El profesional de la información**, 18(5): 493-500.

Orduña-Malea, E.; Serrano-Cobos, J.; Ontalba-Ruipérez, J-A. & Lloret-Romero, N. (2010). Presencia y visibilidad web de las universidades públicas españolas. **Revista española de documentación científica**, 33(2): 246-278.

Ortega, José L.; Aguillo, Isidro F. & Prieto, Jose A. (2006). Longitudinal study of content and elements in the scientific web environment. **Journal of Information Science**, 32(4): 344-351.

Pinto-Molina, M.; Alonso-Berrocal, J. L.; Cordón-García, J. A.; Fernández-Marcial, V.; García-Figuerola, C.; García-Marco, J.; Gómez-Camarero, C.; Zazo, Á. F. & Doucet, A. V. (2004). Análisis cualitativo de la visibilidad de la investigación de las universidades españolas a través de sus páginas web. **Revista española de documentación científica**, 27(3): 345-370.

Pu, H. (2005). A comparative analysis of web image and textual queries. **Online Information Review**, 29(5): 457–467.

Pu, H. (2008). An analysis of failed queries for web image retrieval. **Journal of Information Science**, 34(3): 275–289

Rocki, Marek (2005). Statistical and mathematical aspects of rankings: lessons from Poland. **Higher education in Europe**, 30(2): 173-181.

Sigurbjörnsson, B. & Van Zwol, R. (2008). Flickr tag recommendation based on collective knowledge, Proceedings of the 17th International Conference on World Wide Web 2008, WWW'08, 327-336.

Thelwall, M. & Hasler, L. (2007). Blog search engines. **Online Information Review**, 31(4): 467-479.

Thelwall, M. & Prabowo, R. (2007). Identifying and characterising public science-related concerns from RSS feeds. **Journal of the American Society for Information Science & Technology**, 58(3): 379–390.

Thelwall, M. & Wilkinson, D. (2010). Blog Issue Analysis: An exploratory study of issue-related blogging. In: Birger Larsen, Jesper W. Schneider, Fredrik Åström (Eds.), The Janusz Faced Scholar: A festschrift in honour of Peter Ingwersen, ISSI, 203-218.

Thelwall, M. (2007). Blog searching: The first general-purpose source of retrospective public opinion in the social sciences?. **Online Information Review**, 31(3): 277-289.

Thelwall, M. (2009). Introduction to Webometrics: quantitative web research for the social sciences. San Rafael, CA, Morgan & Claypool, Synthesis Lectures on Information Concepts, Retrieval, and Services, 1(1).

Thelwall, M., & Price, L. (2006). Language evolution and the spread of ideas: A procedure for identifying emergent hybrid word family members. **Journal of the American Society for Information Science and Technology**, 57(10): 1326–1337.

Thelwall, M.; Buckley, K.; Paltoglou, G.; Cai, D. & Kappas, A. (2010). Sentiment strength detection in short informal text. **Journal of the American Society for Information Science and Technology**, 61(12): 2544–2558.

Thelwall, M.; Sud, P. & Vis, F. (2012). Commenting on YouTube videos: From Guatemalan rock to El Big Bang. **Journal of the American Society for Information Science and Technology**, 63: 616–629.

Tolosa, G.; Bordignon, F.; Baeza-Yates, R. & Castillo, C. (2007). Characterization of the Argentinian web. **Cybermetrics**, 11(1): Paper 3. http://cybermetrics.cindoc.csic.es/articles/v11i1p3.html

## Footnotes

[1] <http://www.tineye.com>

[2] <http://www.educacion.es/educacion/universidades/educacion-superior-universitaria/que-estudiar-donde/universidades-espanolas.html> (retrieved on 03/27/2012).
<http://www.crue.org> (retrieved on 03-27-2012).

## Annexes

Annex I. Global graphic file count (Google images)

http://cybermetrics.cindoc.csic.es/articles/v16i1p3_annex1.doc

Annex II. Global graphic file count (Bing images)
http://cybermetrics.cindoc.csic.es/articles/v16i1p3_annex2.doc

Annex III. Global multimedia file count (Google videos)
http://cybermetrics.cindoc.csic.es/articles/v16i1p3_annex3.doc

Annex IV. Global multimedia file count (Bing videos)
http://cybermetrics.cindoc.csic.es/articles/v16i1p3_annex4.doc

Annex V. Global blog count (Google blogs)
http://cybermetrics.cindoc.csic.es/articles/v16i1p3_annex5.doc

**WEBQL.**

TSFA