

The digital press archives of the leading Spanish online newspapers

[Javier Guallar](#) and [Ernest Abadal](#)

Universitat de Barcelona, Facultat de Biblioteconomia i Documentació,
C/ Melcior de Palau, 140, 08014 Barcelona, Spain.

Guallar, J. & Abadal, E. (2010). "The digital press archives of the leading Spanish online newspapers" *Information Research*, **15**(1) paper 424.
[Available at <http://InformationR.net/ir/15-1/paper424.html>]

Abstract

Introduction. This paper analyses the level of development of the press archives of the thirteen Spanish newspapers with the largest digital circulations.

Method. This is evaluative research based on twenty indicators divided into four main sections: general features, the query system, presentation of results and other features. The benchmarks for all the indicators were the *Lexis-Nexis*, *Factiva* and *My News* databases.

Analysis. The press archives of online newspapers were analysed to determine their level of development and the search and browsing facilities they offer according to a set of quality indicators satisfied by professional press databases. The sample consisted of the thirteen newspapers with the largest circulations according to the figures provided by the organisations that audit the Internet press audience. The analysis was carried out from June to August 2007.

Results. The level of development of digital press archives is not homogeneous. Of the newspapers studied, five showed a high or medium-high level of development, though not as high as that of the best commercial products. These were *La Vanguardia*, *ABC*, *El País*, *El Mundo* and *El Periódico*, all of which are online versions of prestigious general information newspapers published in print format. Seven newspapers showed a low or medium-low level of development. These were three daily sports papers (*As*, *Sport*, *Mundo Deportivo*), three exclusively digital newspapers (*Periodista Digital*, *20 Minutos* and *Libertad Digital*) and one regional newspaper (*La Verdad*). Finally, one sports paper (*Marca*) had no search system by keywords.

Conclusions. Fewer than half of the major Spanish newspapers show high or medium levels of development of their digital press archives, though not as high as that of the best press databases, and the rest are still far from satisfying the requirements of a professional search system. The indicators established may also be useful for analysing the situation in other countries.

Introduction

In their short history online newspapers have shown their strong, dynamic nature through constant innovation. Whereas the first online editions of newspapers in the mid-1990s mostly consisted of the print edition with very few changes, the present-day ones may include immediate updating of information, multimedia content and a variety of feedback services.

Access to the content of digital archives is one of the specific new services offered by the digital press. A variety of systems and options are now offered for consulting and viewing back issues. However, the digital press shows a similar attitude to that of the printed press: it considers press archives to be important but not fundamental. Since the 1990s many studies have been made on the way in which online newspapers use the Web to disseminate their contents but few focusing specifically on digital press archives. We will now review several works that have dealt with the subject using a variety of approaches and geographic perspectives.

Cowen described the current situation of British newspapers on the World Wide Web in 2001 and evaluated their future prospects on the basis of interviews with journalists and specialists. She studied the Web sites of five newspapers, and applied sixteen evaluation criteria, including one for digital archives (Cowen [2001](#): 193-194). At that time, all the newspapers analysed allowed their archives (going back one to six years) to be searched free of charge and without registering. Cowen highlighted particularly the content and search facilities of the [Financial Times](#) Web site.

Longo ([2006](#)) reviewed the presence of Italian press archives on the Internet from the viewpoint of a specialised documentalist or researcher. Her non-exhaustive study presents the situation of two main areas: international databases and the media's own press archives. She first analysed the Italian newspapers whose content is marketed by the main international distributors (*Factiva*, *Lexis-Nexis* and *Dialog*). She then described the facilities of the press archives of the main national and regional online newspapers, and reported that digital press archives have started to charge, or at least require registration, for access to back issues.

Among the most important studies of the Spanish press are those by Fuentes, González Quesada and Jiménez López. Their line of research analyses the value-added services of Spanish newspapers that have an online edition, including the consultation of press archives. In two of their first works ([Fuentes and González Quesada 1998](#); [Jiménez López et al. 1999](#)), they carried out an exhaustive study of Spanish newspapers with online versions, analysing indicators such as the coverage of the search system and the operators that it offered. Jiménez López *et al.* ([2000](#)) compared a selection of Spanish press archives with others in the rest of the world and described more than fifty Spanish online newspapers. Just over thirty of them had a digital press archive, but none was comparable to the best of other countries (particularly in the USA), analysed in terms of temporal coverage and search features. The authors also pointed out major differences between the leading Spanish general information newspapers and suggested that these differences could be due to what they considered to be a transitional stage in the formation of digital press archives. Jiménez López ([2003](#)) described the coverage and conditions of subscription to press archives at that time. García Gómez and González Olivares ([2001](#)) studied the search systems of four Spanish online newspapers by making ten searches on current issues and analysing the exhaustiveness and precision of the results. López Carreño ([2004](#)) considered press archives within her categorisation of

digital press services into three groups: information products, document products and value-added services such as access to information adapted to the needs of users. Based on the analysis of online newspapers, she established a newspaper portal model with three levels of service development: basic, intermediate and advanced. She included digital press archives in the basic level of services that should be offered.

Since these studies were carried out, the digital press has obviously continued to evolve, but no research has been done in the last five years to analyse the current or recent situation of Spanish press archives. Only in the annual descriptions of the state of the digital press by Guallar ([2007](#), [2008](#), [2009](#)) is mentioned the perception (from professional use rather than a scientific study) that the search systems of digital newspapers were clearly inferior to those of professional international press archive databases such as *Lexis-Nexis* and *Factiva*, and the Spanish *My News*. For example, searches in the archives of *El País* offered worse results than searches in the *My News* press database. The study stated that '*it would be interesting to have up-to-date scientific data on this (i.e. a comparative study of the real features of the leading Spanish digital media)*' ([Guallar 2007](#): 116). This request was addressed in our study.

Objectives

We set out to determine the level of development and facilities of the digital archives of the leading Spanish online newspapers. We took as a benchmark the most highly developed professional databases in the sector (the international *Lexis-Nexis* and *Factiva*, and the Spanish *My News*), which provide a wide coverage of newspapers from around the world or from Spain. and which also offer among the best search features available in the market, capable of satisfying the most demanding professional needs.

This study can be useful for users (for example, journalists, librarians, researchers and students) who use digital press archives for their work and also for developers of this kind of portal. Our analysis shows the level of development of leading Spanish newspapers and this can be useful in helping users decide which press archive would be better for their needs, as well as helping developers improve the features of their product.

Methods

The content of digital press archives was analysed to determine their degree of compliance with a set of indicators. Our study is thus different from those carried out by Lin and Jeffres ([2001](#)), who performed a content analysis of 422 Web sites of newspapers, radio stations and television stations in the largest metropolitan areas of the USA, or Stryker *et al.* ([2006](#)), who analysed the treatment of cancer in the press through searches of press databases. These studies required mechanisms to ensure a high degree of reliability among the analysts, to achieve good rates of coincidence, and so on. In our case, on the other hand, it was only necessary to determine the presence or absence of indicators on which both authors performed the analysis and agreed the results.

The sample consisted of the thirteen newspapers with the largest circulations according to the figures provided by the Spanish organisations that audit Internet press audience. The analysis was carried out from June to August 2007.

Choice of the samples

The two most important systems for measuring press circulation and audience in Spain are the [Estudio General de Medios](#) (General Media Study) and the [Oficina de Justificación de la Difusión](#) (Circulation Justification Office).

The *Estudio General de Medios* is an audience-measurement system based on surveys of users. The twenty-five most visited Spanish Web sites in February-March 2007 included eleven online newspapers: *Marca*, *El País*, *As*, *Mundo Deportivo*, *Sport*, *La Vanguardia*, *ABC*, *El Periódico*, *El Correo*, *Expansión* and *Norte de Castilla*. However, this list does not include some online newspapers, such as the newspaper *El Mundo*, because they do not accept its measurement system.

The *Oficina de Justificación de la Difusión* measurement system is based on the visits to the Web sites of online newspapers. According to the figures it published for March 2007, the following are the twenty Spanish online newspapers with the largest circulations: *El Mundo*, *Marca*, *20 Minutos*, *ABC*, *Periodista Digital*, *Libertad Digital*, *Sport*, *La Verdad*, *Ideal*, *El Correo*, *El Periódico de Catalunya*, *Las Provincias*, *El Confidencial*, *La Voz de Galicia*, *El Comercio*, *La Razón*, *La Nueva España*, *Expansión*, *Hoy-Diario de Extremadura* and *Europa Press*. In this case, the media that do not accept its control are the newspapers *El País*, *As*, *La Vanguardia* and *Mundo Deportivo*.

To obtain a sample that was as representative as possible and did not leave out any of the most important newspapers (other than those that exclude themselves voluntarily from both lists), it was decided to combine the eight first results of each list. This gave a sample of thirteen digital media: *Marca*, *ABC* and *Sport* are common to both lists; *El País*, *As*, *Mundo Deportivo*, *La Vanguardia* and *El Periódico* are only in the *Estudio General de Medios* list; and *El Mundo*, *20 Minutos*, *Periodista Digital*, *Libertad Digital* and *La Verdad* are only in the *Oficina de Justificación de la Difusión* list.

The sample is sufficiently varied because it includes general and sports press, national and regional press, paid and free press, and exclusively digital and digital-print press.

Name	Subject coverage	Geographic coverage	Paid/Free press	Format
20 Minutos	general	national	free	digital-print
ABC	general	national	paid	digital-print
As	sport	national	paid	digital-print
Libertad Digital	general	national	free	digital
Marca	sport	national	paid	digital-print
El Mundo	general	national	paid	digital-print
Mundo Deportivo	sport	national	paid	digital-print
El País	general	national	paid	digital-print
El Periódico	general	national	paid	digital-print

Periodista Digital	general	national	free	digital
Sport	sport	national	paid	digital-print
La Vanguardia	general	national	paid	digital-print
La Verdad	general	regional	paid	digital-print

Table 1: The sample of newspapers analysed

Indicators for the evaluation

Digital press archives are databases. In this field, we can highlight four texts: two dealt with the process of consulting databases from the viewpoint of the process followed by the user ([Marchionini 1995](#) and [Shneiderman et al. 1997](#)), and another two presented the characteristics of a good query interface ([Nielsen and Loranger 2006](#) and [Morville and Rosenfeld 2006: 145-192](#)). Based on these studies, an original proposal was made of the fundamental elements for database query interfaces ([Abadal 2002](#)), which has been taken into account for the present study.

Our evaluation proposal is based on twenty indicators divided into four main sections: general features, the query system, presentation of results, and other features. Table 2 includes a brief description of each indicator and the way it is evaluated.

The two most important features for users in a database are the search facilities and the presentation of results. Because of that, we have established a higher number of indicators for these. Although we have considered these features as critical points for the evaluation of databases, we have also included other indicators to complete the evaluation: for example, temporal coverage (another critical issue) and others that describe press archives without evaluating them (indicators 1.1, 1.2, 4.1 and 4.3 in Table 2).

Section	Indicator	Description
General features	1.1. Name and location	The name that the newspaper gives to its search system or systems, and its location on the home page. This indicator was not used for the evaluation.
	1.2. Information on the technology used	Information and explanations on the program used and its accountability (whether it is an in-house or external system). This indicator was not used for the evaluation.
	1.3. Coverage	The temporal coverage of the press archive. The proportion of the archive that is accessible online and whether it had a coverage of at least five years.

Query system	2.1. Types of search	Types of query by keyword available. Whether the press archive had more than one search option, such as simple and advanced.
	2.2. Combination of search terms	Advanced features for search terms: Boolean operators, search for literal expressions, proximity operators, use of parentheses to increase combinations, and adjusting the relevance percentage of the query. In order to fully satisfy this indicator, the system must offer at least searches with Boolean operators and literal expressions.
	2.3. Search by date	Facilities for search by date. It considers whether the system allows the exact dates of the search to be defined.
	2.4. Search by collection	Ability to differentiate between global and partial searches of the collections in the different sections of the newspaper, supplements, etc.
	2.5. Search by document field	Ability to differentiate between global or partial searches in fields such as title, author and section.
	2.6. Reusing search strategies	Facilities for saving previous queries, if the system has a search history option.
	2.7. Search by browsing	Possibility of access to the document by browsing issues of the newspapers.
	2.8. Help	Existence of explanatory texts on the use of the search system.
Presentation of results	3.1. Management of the results lists	Options for managing the results, such as ordering them by relevance and date or limiting the number of results per page.
	3.2 Document fields	Number and type of fields shown in each result (e.g. author, title, date, section, etc.). A search engine is considered to satisfy this indicator if it offers at least four fields.
	3.3. Identification of the search terms in the	Highlighting of the search terms in the results or in the document.

	document	
	3.4. Choice of formats for viewing the document	Different document formats (html, pdf, etc.).
	3.5. Options for managing the documents obtained	Different ways of managing the documents obtained: sending by email, printing, saving, obtain use statistics, evaluating or commenting on the news item, sharing it (sending it to social Web sites), etc. A search engine is considered to satisfy this indicator if it offers at least four options.
	3.6. Presentation of related documents	Obtaining documents related to the results in the same press archive or external sources.
Other features	4.1. Accessibility	Degree of compliance with accessibility standards that allow disabled persons to consult Web sites. For the quantitative analysis the <i>Tawdis</i> program was used. This indicator was not used for the evaluation.
	4.2. Visibility	The impact of the press archive page is evaluated on the basis of the number of links to it from Web pages on other domains. The <i>Yahoo Site Explorer</i> service was used to count them.
	4.3. Cost	Free or paid access to the press archive. This indicator was not used for the evaluation.
Table 2: Indicators for analysing digital press files		

The benchmarks for all the indicators were the *Lexis-Nexis*, *Factiva* and *My News* databases (see above). The unit of analysis was the set of pages of the press archive section of the newspaper.

Analysis

General features

Name and location

The terms *archivo* (archive) and *hemeroteca* (newspaper library) are both used (in four and six cases, respectively). In fact, *ABC* and *El Periódico* use both names to differentiate the two different search systems that they offer on the same Website. In

cases in which there is no specific section, the term *search* or *search engine* is normally used to inform users that they can consult back issues.

Information on the technology used

The newspapers offer little or no information on the technology used to manage their digital press archives. *ABC*, *El Periódico* and *Sport* all offer two search systems on their Web sites, one of their own, on which they provide no information, and one contracted to the well-known Spanish press database *My News*. *El Mundo* uses the *Autonomy* search engine, version beta v2, and *La Verdad* uses the *Sarenet* search system. *Libertad Digital*, *Periodista Digital* and *20 Minutos* use the *Google* and *Yahoo!* search engines, linking to themselves. They state, sometimes very explicitly, that their Web sites decline all responsibility for the results of the search system.

Coverage

Spanish newspapers have increased the scope of their digital archives, but still do not offer the whole of their collections, with the important exceptions of *El País* and *La Vanguardia*, which have placed online the whole of their print archive since their foundation in 1976 and 1881, respectively. *ABC* offers its collection from 1996 and *El Mundo* from 1994. *El País* presents its whole print archive in html, whereas *La Vanguardia* presents it in pdf files. The newspapers mostly offer a shorter period of coverage for the archives of their digital versions (Table 3), which go back between one and seven years.

Newspaper	Print edition	Digital edition
<i>20 Minutos</i>	January 2005 (pdf)	January 2005
<i>ABC</i>	June 1996 (html) / Last 15 days (pdf)	January 2002
<i>As</i>	July 2004 (html) / January 2004 (pdf)	July 2004
<i>Libertad Digital</i>	n/a	No information available
<i>Marca</i>	No information available	No information available
<i>El Mundo</i>	January 1994 (html) / October 2002 (pdf)	January 2000
<i>Mundo Deportivo</i>	January 2004 (html) / Last 15 days (front page) (pdf)	January 2004
<i>El País</i>	4 May 1976 (html) / 25 July 2001 (pdf)	No information available
<i>El Periódico</i>	January 2000 (html, pdf)	January 2006
<i>Periodista Digital</i>	n/a	September 2006
<i>Sport</i>	January 2005 (pdf)	January 2006
<i>La Vanguardia</i>	January 1999 (html) /	November 2000

	February 1881 (pdf)	
<i>La Verdad</i>	January 2006 (html)	January 2006

Table 3: Temporal coverage of Spanish online newspapers

The query system

Types of search

Nine of the digital press archives offer two types of search, simple and advanced, though the names used to refer to them vary. Three others offer a single option, in *La Vanguardia* an advanced search and *20 Minutos* and *Libertad Digital* a simple *Google* and *Yahoo!* search, respectively. *ABC*, *El Periódico* and *Sport* each have two search systems on their Web sites: a *My News* search and a proprietary one. Finally, *Marca* is the only digital press archive that has no search engine: users must browse instead.

Combination of search terms

All the retrieval systems use Boolean operators, either shown directly (AND, OR, NOT) or masked through expressions (search for any words, all words, exact phrase, etc.). However, they lack some of the more sophisticated functions offered by professional press archives, such as using parentheses to increase the combination of terms or adjusting the relevance of the results. The only exception to this is *El Mundo*, which offers both of these options.

Search by date

Six of the digital press archives allow users to choose the search dates through a drop-down menu with closed options (today, last week, last month, etc.) or between two exact dates. Two of them only offer the possibility of searching between two dates (*La Vanguardia* and *Mundo Deportivo*), one offers a drop-down menu (*El Mundo*), and two do not have this option at all (*20 Minutos* and *Libertad Digital*).

Search by collection

Ten of the digital press archives offer the possibility of restricting the search to a given collection (print, digital, supplements, etc.), though the scope of this function varies. *El País*, *La Vanguardia* and *ABC* offer a large number of collections. For example, *El País* offers: print edition with its sections, supplements and regional editions (Madrid, Cataluña, etc.), digital edition with its sections, other media from *Prisa* group (*Cadena Ser*, *Cinco Días*, etc.), and search by multimedia collections (photographs, videos, etc.). *El Mundo*, *La Verdad* and *El Periódico*, offer a medium number; and the rest offer few collections or do not have this option at all (*20 Minutos* and *Sport*).

Search by document field

This option, which is common in professional databases, is only found in four of the thirteen systems analysed: *As*, *Mundo Deportivo*, *El País* and *La Vanguardia*. It is noteworthy that some of the best query systems, such as those of *El Mundo* and *ABC*,

do not have this function. This is a very important shortcoming in which most of the newspapers show room for improvement.

Reusing search strategies

None of the digital press archives has search history options for reusing previous search strategies.

Search by browsing

All the digital press archives offer access through browsing, generally by means of monthly calendars (including *Marca*, for which it is the only system). Originally the only system in many cases, this form of access is now complementary to searches by keywords.

Help

Only six digital press archives provide instructions for the search function, and in some cases these are very simple. It is surprising that such a basic element for users of a query system has been neglected by half the newspapers studied, particularly *El País*, which is among the leaders in most indicators. This is one of the main shortcomings of the systems analysed.

General evaluation

The query systems of the digital press archives analysed are at an intermediate level of development in comparison with the best commercial databases. They include the main Boolean operators but lack the more sophisticated options such as the use of parentheses to increase the combination of terms and, with the sole exception of *El Mundo*, they do not allow the relevance to be adjusted. They offer good facilities for searching by date and by collections. However, features such as searching by document field and reusing search strategies through search histories, which are common in professional bibliographic databases, are almost non-existent in the main online newspapers. Finally, one of the main shortcomings is that half the archives in the sample lack help services, which are basic for users of query systems.

Presentation of results

Management of the results lists

Seven of the online newspapers offer two functions that can be considered as basic in a professional query system: ordering of the results by relevance or by date, and selection of the number of hits displayed on each page. These are: *ABC*, *As*, *El Mundo*, *El País*, *El Periódico*, *Sport* and *La Verdad* (though, as stated above, *El Periódico* and *Sport* have two different search systems, and one of them does not offer this option). The other five, *20 Minutos*, *Libertad Digital*, *Mundo Deportivo*, *Periodista Digital* and *La Vanguardia*, offer no results management system. This group includes the two digital newspapers and the free one, which in general show lower values in other indicators.

Document fields

Most of the digital press archives analysed show between four and six document fields, though *ABC* shows eight fields in the Archivo de *ABC* system, and *El Periódico* shows only two in one of its query systems (but six in the other one). Except in one case, one of the fields is the headline, with a hyperlink to the news item.

Identification of the search terms in the document

This indicator shows low values overall, because half the sample do not have this feature. The other half highlight the search term in the lines of text displayed, and in one case also in the headline. The effectiveness of this feature depends on the amount of text shown in the result: four newspapers show only one or two lines of text and only two (*El País* and *La Vanguardia*) show four and six lines, thus taking full advantage of this feature.

Choice of formats for viewing the document

Eight of the online newspapers with independent print and digital editions show their information in both the html format of the Web site and the pdf format of the print edition. However, two media with print and digital editions offer only html: *Mundo Deportivo* (which offers only the front page of the print edition as a jpg image) and *La Verdad*. The two exclusively digital newspapers only present the information in html. The pdf format, therefore, is the standard for presentation of the print edition of the newspapers. Furthermore, *El Mundo* and *El País* offer specific versions in accessible text (or plain text) and for PDAs and mobile phones.

Options for managing the documents obtained

Two elementary document management options are very common: printing the news item and sending it by e-mail, which are found in eleven and ten of the digital press archives, respectively. The participatory services of a) voting and/or commenting on the news item, and b) sending it to social Web sites (*Digg*, *Meneame*, *Delicious*, *My Yahoo*, *Technorati*) are offered by about half the newspapers analysed (six and five, respectively). In this case, the most advanced media in Web 2.0 services are the free newspaper *20 Minutos*, the exclusively digital *Periodista Digital* (both have shown a concern for this type of service for some time) and the traditional newspapers *ABC*, *El País* and *La Vanguardia*. However, none of them allow part of the result to be selected for further operations (print, send, etc.), which must be performed individually, record by record.

Presentation of related documents

Eight press archives give access to a service of great interest: showing news related to the results. In six cases they offer only news from their own archives, but *El País* and *Libertad digital* go one step further by linking to related news in other media. Four do not offer this feature (*As*, *El Periódico*, *Sport* and *La Verdad*). Linking to external Web sites, which is a common practice in other contexts such as blogs, is not common in the main online newspapers, due to their competing position in the market

General evaluation of the presentation of results

Only just over half the media consulted allow users to order the results by relevance and date, and to determine the number of results presented on each page. The press archives show on average between four and six fields for the document record. The fields offered in all cases were the date, the headline with a link to the news item, and a few lines of text (mostly short). Half of them highlight the search terms in the document record, and of these only two (*El País* and *La Vanguardia*) take full advantage of it. All offer the texts in the html format of the Web and eight of them (the newspapers with a print version) also offer texts in pdf, which is thus consolidated as the standard for disseminating printed texts on the Internet. Two options for managing the documents obtained were offered by the vast majority: printing and sending by e-mail. The recently introduced service of sending them to social Web sites like *Digg*, *Technorati* and *Delicious* is becoming increasingly widespread, but is still offered by only half the newspapers. Finally, the presentation of news related to the results is widespread, but only *El País* and *Libertad digital* offer links to other newspapers outside their own press archive.

Other features

Accessibility

The degree of accessibility of a Web page is measured by the ease with which any type of user can access its contents. The World Wide Web Consortium (W3C) has published a series of guidelines that allow persons with disabilities to access Web page content. These guidelines establish the requirements that must be fulfilled, for example, by animations and images (which are described with the ALT attribute) or hypertext links (the linked expression should have meaning outside the context: 'Click on the W3C report', rather than 'Click here'). Furthermore, they ensure accessibility regardless of the users' computer, screen, browser and type of connection. For the study we used the [TAW \(Web Accessibility Test\)](#), a tool for the analysis of Web sites based on the W3C - Web Content Accessibility Guidelines 1.0 (WCAG 1.0), which detects three types of accessibility errors for Web pages. Failure to comply with priority 1 means that one or more groups of users will find it impossible to access information in the document; failure to comply with priority 2 means that they will find it difficult to access information in the document; and failure to comply with priority 3 means that they will find it somewhat difficult to access information in the document.

The following table shows the results of the test. They are in general fairly poor, showing a lack of concern for this subject by the developers of the Web sites of online newspapers. The two values indicated for each priority refer to the number of manual and automatic errors.

Newspaper	P1	P2	P3
20 Minutos	15-155	10-354	10-75
ABC	0-159	564-373	12-37
As	67-243	235-455	16-97
Libertad Digital	23-118	35-197	2-17
Marca	0-74	43-326	4-30

El Mundo	10-90	67-59	18-36
Mundo Deportivo	9-214	55-311	10-26
El País	3-98	230-129	19-28
El Periódico	6-320	89-621	1-39
Periodista Digital	14-223	54-305	4-67
Sport	3-151	44-242	1-24
La Vanguardia	2-96	27-107	5-37
La Verdad	6-79	30-150	17-59

Table 4: Compliance with W3C accessibility guidelines

Visibility

The aim of this indicator is to analyse the presence of the press archives on the World Wide Web based on the number of specific links to their Web sites.

Gao and Vaughan (2005) analysed the visibility of four newspapers of the USA, Canada, China and Hong Kong from a study of a sample of links to their Web sites. As has been done in similar studies, they differentiated between internal links (from the same Web site) and external ones. They tested different search engines (*Google*, *MSN* and *Yahoo!*) and chose *Yahoo!* because it had a greater number of links and allowed the external links to be distinguished from the total.

We decided to follow the same methodology and used the new [Yahoo Site Explorer](#) service, which offers many facilities for finding only internal links. In two cases, however, the number of links was increased considerably by links from the same publishing group.

The following table shows the number of direct links to the Web page of the newspaper and to the Web page of the press archive.

As can be seen, the number of links to the press archive pages is low, particularly compared with the number of links to the home page of the newspaper. *El Mundo* stands out with a very high number of links, followed by *ABC*. The extremely low number of links of *El País* is related to the change of domain from *elpais.es* to *elpais.com*.

Newspaper	Links to the archive	Links to the newspaper
20 Minutos	1,705	262,615
ABC	5,620	323,685
As	-	825,359
Libertad Digital	-	128,645
Marca	-	188,199
El Mundo	14,307	747,534

Mundo Deportivo	-	161,852
El País	734	518,984
El Periódico	1,290	1,377,886
Periodista Digital	-	139,591
Sport	-	1,306,049
La Vanguardia	396	134,630
La Verdad	14	590

Table 5: Number of links to the press archive and to the newspaper

Cost

Though *El País* charged for all its content between November 2002 and June 2005 ([Guallar 2007](#)), there are currently no newspapers in Spain that do so. Eight of the newspapers follow the mixed model (free and paid content), which consists in offering the online edition free and charging for part or all of the print edition. The normal systems of payment include annual and half-year subscription and the possibility of purchasing news and back issues separately, generally with discounts for packages. The price of the subscription ranges between €75 and €9 a year and €50 and €60 for a half-year, with the exception of *As* (€235 and €130, respectively) and *Marca* (€20 per month). There are a variety of package options.

Five online newspapers are completely free: *Periodista Digital*, *Libertad Digital*, *20 Minutos*, *La Verdad* and *Mundo Deportivo*. All of these have two characteristics: they do not have content in pdf format, and they generally have lower scores than the pay newspapers in the quality indicators of their search systems.

Newspaper	Annual subscription	Purchase of single issues
20 Minutos	Free	
ABC	Annual €80, half-year €50	€0.75 news item, packages
As	Annual €235, half-year €130	€1.90, day €0.90
Libertad Digital	Free	
Marca	Monthly €20	€0.60 per issue
El Mundo	Annual €75 (only pdf), €97 (html and pdf)	Packages of articles and issues; e.g. 100 articles, €9
Mundo Deportivo	Free	
El País	€80	€0.50 per issue
El Periódico	No	€0.50 article (html), €0.75 (html and pdf),

		with packages
Periodista Digital	Free	
Sport	Annual €80, half-year €50	pdf news item €2 with packages
La Vanguardia	Annual €95, half-year €60	€0.9 per issue, €3 per issue from the historic archive
La Verdad	Free	

Table 6: Cost and form of payment

Results

The level of development of the press archives in the Spanish newspapers with the widest circulations is not homogeneous. Table 7 presents a summary of the results of the press archives.

Newspaper	Indicators																Total
	1.3	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	3.1	3.2	3.3	3.4	3.5	3.6	4.2	
20 Minutos								X			X	X		X	X	X	6
ABC	X	X	X	X	X			X	X	X	X		X	X	X	X	13
As		X		X		X		X		X	X		X	X			8
Libertad Digital			X		X				X		X	X			X		6
Marca								X									1
El Mundo	X	X	X		X			X	X	X	X		X	X	X	X	12
Mundo Deportivo		X	X	X	X	X					X				X		7
El País	X	X		X	X	X		X		X	X	X	X	X	X	X	13
El Periódico	X	X		X	X	X			X	X	X	X		X	X	X	12
Periodista Digital		X	X		X			X			X	X		X	X		8
Sport		X		X				X	X	X	X		X	X			8
La Vanguardia	X	X	X	X	X	X		X	X		X	X	X	X	X	X	14
La Verdad		X	X	X	X			X		X	X						7
Total	5	10	8	8	9	4	0	11	6	7	12	5	7	9	8	6	

Table 7: Compliance of press archives with the indicators (see Table 2 for indicators)

A quantitative analysis of the results was used to draw up the following ranking of newspapers (from a total of sixteen):

- *La Vanguardia*: 14
- *ABC, El País*: 13
- *El Mundo, El Periódico*: 12
- *As, Periodista Digital, Sport*: 8
- *Mundo Deportivo, La Verdad*: 7
- *20 Minutos, Libertad Digital*: 6
- *Marca*: 1

Three main groups of newspapers can be distinguished.

1. Medium-high to high

Five newspapers comply with twelve to fourteen of the sixteen indicators evaluated. They are all digital versions of prestigious general newspapers published in print format: *La Vanguardia*, *Abc*, *El País*, *El Mundo* and *El Periódico*. First and second place are occupied by *La Vanguardia* and *El País*, the only newspapers that offer their whole archive digitally (over a hundred years for the former). *El Mundo* has the best features for combining search terms, which is the most important indicator for a query system. Despite their high score, *ABC* and *El Periódico* have press archives with two unequal search systems, as stated above.

2. Medium-low

Seven newspapers comply with six to eight of the indicators. These are three daily sports papers (*As*, *Sport* and *Mundo Deportivo*), three exclusively digital newspapers (*Periodista Digital*, *20 Minutos* and *Libertad Digital*) and a regional newspaper (*La Verdad*). This group is far behind the first group. Though they show high values in some points, on the whole the level of development of their press archives is limited and they show much room for improvement.

3. Low

The sports paper *Marca* only complies with one indicator (that of searching by browsing) because it has no search engine. Despite the wide audience of this publication, it has not yet decided to offer a keyword search system.

Conclusions

In summary, fewer than half of the major Spanish newspapers show high or medium levels of development of their digital press archives, though not as high as that of the best press databases, and the rest are still far from satisfying the requirements of a professional search system. By types of media, the sports papers lag behind the general information newspapers in the quality of their search systems, exclusively digital newspapers with more recent press archives fail to give them sufficient importance, and regional newspapers lag behind the major dailies.

The indicators that obtained the highest scores were search by browsing (twelve newspapers) and document fields (eleven newspapers). Those that obtained the lowest scores were repeat search functions (offered by no newspapers), partial searches in results fields (four newspapers), highlighting of search terms (five newspapers), temporal coverage (still low except in five newspapers), visibility (six newspapers) and help texts (six newspapers). Thus, these are the areas in which Spanish digital press archives should improve their services in the future.

As a result of this analysis, several recommendations can be made for developers and publishers:

Enhance contents

The most desirable thing is to make the entire newspaper available digitally, including supplements. And, of course, the original digital version should be available too, since it is sometimes different to the print one. Nevertheless, at the time of our study, very few newspapers actually make the entire publication available. Newspapers that offer only a reduced portion of their content on the Web are missing out on the enormous potential of a large newspaper archive because this is a good way to attract new visitors and to maintain the interest of subscribers.

More search capacity

Our study clearly identifies the existence of two groups of newspapers. Those that are more advanced would have only to slightly improve their search services, whereas the less advanced group would need to introduce major improvements. Regardless of the group, newspapers in each should pay special attention to those indicators that received the lowest scores. The existence of a professional-level information retrieval system could substantially improve access to contents of high interest for a wide variety of users.

We have not focussed any attention on user studies in this text, but they are certainly an important tool that could be applied here. The best way for improving the quality and utility of newspaper archives is to analyse their use, through methods such as log analyses, eye-tracking studies and surveys.

Finally, our study can be useful for both professional and non-professional users by allowing them to identify the best newspaper archives and by helping them to choose the best options for consulting retrospective news.

Note

This paper bears some relationship to a text by the same authors ([Guallar and Abadal 2009](#)) published in Spanish in *El profesional de la información*. The objective of this earlier article was to establish a list of indicators for the evaluation of digital press archives and to present examples of good practices. The current text applies these same indicators to the thirteen Spanish newspapers with the largest digital circulation, presents the quantitative analysis of the results and ranks the newspapers according to the level of development of their press archives.

About the authors

Javier Guallar is Professor in the Department of Library and Information Science, University of Barcelona. He is vice-director of [*El Profesional de la Información*](#). He can be contacted at: jguallar@gmail.com

Ernest Abadal is Senior Lecturer in the Department of Library and Information Science, University of Barcelona. He is director of the scientific journal [*BiD: textos universitaris de biblioteconomia i documentació*](#). He can be contacted at: abadal@ub.edu

References

- Abadal, E. (2002). [Elementos para la evaluación de interfaces de consulta de bases de datos](#) [Elements for an evaluation of database interfaces]. *El Profesional de la Información*, **11**(5), 349-360. Retrieved 2 September, 2009 from <http://www.elprofesionaldelainformacion.com/contenidos/2002/septiembre/3.pdf> (Archived by WebCite® at <http://www.webcitation.org/5jFMhhmbm>)
- Cowen, N. (2001). The future of the British broadsheet newspaper on the World Wide Web. *Aslib Proceedings*, **53**(5), 189-200.
- Fuentes, M.E. & González Quesada, A. (1998). La prensa española en Internet: análisis de los servicios de valor añadido [The Spanish press on the Internet: analysis of value added services]. In *VI Jornadas Españolas de Documentación Automatizada, Valencia, 1998*. (pp. 281-292). Valencia: Federación Española de Sociedades de Archivística, Biblioteconomía y Documentación.
- Gao, Y. & Vaughan, L. (2005). Web hyperlink profiles of news sites: a comparison of newspapers of USA, Canada, and China. *Aslib Proceedings*, **57**(5), 398-411.
- García Gómez, J.C. & González Olivares, J.L. (2001). Aproximación a la evaluación cuantitativa de los sistemas de recuperación de información de la prensa en Internet: exhaustividad y precisión [Approach to quantitative evaluation of information retrieval systems of the press on the Internet: completeness and accuracy]. *Scire*, **7**(1), 143-152.
- Grau, J. & Guallar, J. (2004). [My News, la hemeroteca digital de la prensa española](#) [My News, the digital press archive of the Spanish press]. *El Profesional de la Información*, **13**(6), 466-476. Retrieved 2 September, 2009 from <http://eprints.rclis.org/archive/00008532/01/My-News.pdf> (Archived by WebCite® at <http://www.webcitation.org/5k9FHdsFR>)
- Guallar, Javier (2003). [Mètodes i tècniques de recerca en els articles de documentació periodística a Espanya](#). *BiD: Textos Universitaris de Biblioteconomia i Documentació*, No. 11. Retrieved 2 September, 2009 from http://www2.ub.es/bid/consulta_articulos.php?fichero=11gualla.htm (Archived by WebCite® at <http://www.webcitation.org/519lv1QKS>)
- Guallar, J. (2007). [Prensa digital en 2006](#). [Digital journalism in 2006]. In *Anuario ThinkEpi 2007* (pp. 106-117). Barcelona: EPI SCP. Retrieved 2 September, 2009 from <http://eprints.rclis.org/archive/00008388/01/AnuarioThinkEPI2007-Guallar-Prensa-digital-en-2006.pdf> (Archived by WebCite® at <http://www.webcitation.org/5k9FNaZAw>)

- Guallar, J. (2008). Prensa digital en 2007. [Digital journalism in 2007]. In *Anuario ThinkEpi 2008* (pp. 102-108). Barcelona: EPI SCP.
- Guallar, J. (2009). Prensa digital en 2008. [Digital journalism in 2008]. In *Anuario ThinkEpi 2009* (pp. 88-94). Barcelona: EPI SCP.
- Guallar, J. & Abadal, E. (2009). [Evaluación de hemerotecas de prensa digital: indicadores y ejemplos de buenas prácticas](#) [Evaluation of digital press archives: indicators and examples of good practice]. *El Profesional de la Información*, **18**(3), 255-269. Retrieved 2 September, 2009 from http://eprints.rclis.org/16899/1/epi09_guallar-abadal_evaluacion_hemerotecas.pdf (Archived by WebCite® at <http://www.webcitation.org/5l9dqKvqH>)
- Jiménez López, Á. (2003). [Hemerotecas de pago en la prensa digital española](#) [Payment press archives in the Spanish digital press]. *El Profesional de la Información*, **12**(6), 473-474. Retrieved 2 September, 2009 from <http://www.elprofesionaldelainformacion.com/contenidos/2003/noviembre/9.pdf> (Archived by WebCite® at <http://www.webcitation.org/5k9Hr4VhI>)
- Jiménez López, Á. (2002). [Estudi de la gestió documental de la informació en els serveis de valor afegit dels mitjans de comunicació a Internet: el cas de la premsa diària a l'Estat espanyol](#) [Study of information management on value added services from Internet media: the case of newspapers in Spain]. Bellaterra: Universitat Autònoma de Barcelona. Retrieved 2 September, 2009 from <http://www.tesisenxarxa.net/TDX-1008102-143437/index.html> (Archived by WebCite® at <http://www.webcitation.org/5k9HtaUCs>)
- Jiménez López, Á., González Quesada, A. & Fuentes, M. E. (2000). [Las hemerotecas digitales de la prensa en Internet](#) [Digital newspapers libraries on the Internet]. *El Profesional de la Información*, **9**(5), 15-24. Retrieved 2 September, 2009 from <http://www.elprofesionaldelainformacion.com/contenidos/2000/mayo/2.pdf> (Archived by WebCite® at <http://www.webcitation.org/5k9Hwa5HI>)
- Jiménez López, Á., González Quesada, A. & Fuentes, M.E. (1999). [Gestió documental de la informació en els serveis de valor afegit de la premsa espanyola a internet](#) [Information management on value added services from Spanish press on Internet]. In *7es Jornades Catalanes de Documentació* (pp. 405-417). Barcelona: Col·legi Oficial de Bibliotecaris-Documentalistes de Catalunya. Retrieved 2 September, 2009 from <http://www.cobdc.org/jornades/7JCD/18.pdf> (Archived by WebCite® at <http://www.webcitation.org/5k9HygEmq>)
- Lin, C.A. & Jeffres, L.W. (2001). Comparing distinctions and similarities across Web sites of newspapers, radio stations, and television stations. *J&MC Quarterly*, **7**(3), 555-573.
- Longo, B. (2006). Gli archivi dei giornali online [The archives of online newspapers]. *Biblioteche Oggi*, **24**(1), 9-21.
- López Carreño, R. (2004). [Análisis taxonómico de los portales periodísticos españoles](#) [Taxonomic analysis of Spanish newspaper Web sites]. *Anales de Documentación*, **7**, 123-140. Retrieved 2 September, 2009 from <http://www.um.es/fccd/anales/ad07/ad0708.pdf> (Archived by WebCite® at <http://www.webcitation.org/5k9I1SJeI>)
- Marchionini, G. (1995). *Information seeking in electronic environments*. Cambridge: Cambridge University Press.

- Morville, P. & Rosenfeld, L. (2006). *Information architecture for the World Wide Web* (3rd. ed.). Sebastopol, CA: O'Reilly.
- Nielsen, J. & Loranger, H. (2006). *Prioritizing web usability*. Berkeley, CA: New Riders.
- Neilson, K & Willett, P. (1999). United Kingdom regional newspapers on the World Wide Web. *Aslib Proceedings*, **51**(3), 78-90
- Shneiderman, B., Byrd, D. & Croft, W.B. (1997, January). [Clarifying search: a user-interface framework for text searches](#). *D-Lib Magazine*. Retrieved 2 September, 2009 from <http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html> (Archived by WebCite® at <http://www.webcitation.org/5k9I40g6V>)
- Stryker, J.E., Wray, R., Hornik, R.C. & Yanovitzky, I. (2006). Validation of database search terms for content analysis: the case of cancer news coverage. *J&MC Quarterly*, **83**(2), 413-430.

How to cite this paper

Guallar, J. & Abadal, E. (2010). "The digital press archives of the leading Spanish online newspapers" *Information Research*, **15**(1) paper 424. [Available at <http://InformationR.net/ir/15-1/paper424.html>]

© the authors, 2010.

Last updated: 15 February, 2010