Search Engine Optimization and Online Journalism: 
*The SEO-WCP Framework*

Lluís Codina, Mar Iglesias-García, Rafael Pedraza & Lucía García-Carretero

A DigiDoc – UPF Research Group Publication
April 2016
**Serie Editorial DigiDoc**

A deliverable in the *Active Audiences and Journalism* collection


**NS PAA04/2016**

A production of

[Universitat Pompeu Fabra](https://www.upf.edu)  
[Departament de Comunicació](https://www.upf.edu)

With the support of
Executive summary

The dissemination of news stories today takes place via various platforms, among which news media websites are just one. In other words, the audience for journalistic content, that is, those people who wish to be informed about the world around them, access news stories via the results pages of search engines as well as via social networks, and not solely via the websites of the news media.

This means the news media today have to implement an effective search engine optimization (SEO) policy to ensure their success, otherwise, these platforms (the results page and social networks) will not provide the percentage of visibility and visits that these online news sites should be obtaining.

This document presents a framework for optimizing journalistic content, both from the perspective of web optimization or SEO and from the perspective of visibility on social networks. The framework can be characterized as being:

- A methodology for training and for the acquisition of skills in the field of SEO and Communication.
- A proposal for independent work, but one that is at the same time adaptable and compatible with the style books of different news media.
- A proposal for new cyber media that need to adopt a methodology to begin their professional activities.
- A proposal for comparing standards and procedures that any researcher or firm can adopt as part of their comparative analyses and benchmarking studies to improve procedures.
About the authors

Lluís Codina is an Associate Professor at the Pompeu Fabra University (UPF) in Barcelona. He teaches in the Faculty of Communication on the undergraduate courses in Journalism and Audiovisual Communication. He is the coordinator of the UPF Master’s in Social Communication in the Department of Communication. He also forms part of the academic staff that teaches on the Online Master’s in Digital Documentation and Search Engines organized by the UPF’s Institute of Continuing Education. He is an associate member of the Digital Documentation and Interactive Communication (DigiDoc) Research Group and coordinator of its Research Seminars. ORCID: https://orcid.org/0000-0001-7020-1631 Contact: lluis.codina@upf.edu

Mar Iglesias-García is a journalist and lecturer in the Department of Communication and Social Psychology at the University of Alicante. She teaches on the undergraduate course in Advertising and Public Relations and on the course in Tourism. She also participates in the research project entitled “Active audiences and journalism. Interactivity, web integration and searchability of journalistic information”, funded by the Ministry of Economy and Competitiveness. Since 2010, she has been the director of the cyber newspaper ComunicandoUA. ORCID: https://orcid.org/0000-0001-7926-5746 Contact: mar.iglesias@gcloud.ua.es

Rafael Pedraza-Jiménez is an Associate Professor in the Department of Communication at the Pompeu Fabra University (UPF). He teaches on the undergraduate courses in Journalism and Audiovisual Communication. He also teaches on various masters’ and postgraduate courses at the UPF and at other universities. As a researcher, he coordinates the DigiDoc Research Group. Throughout his career, he has participated and/or led a number of public and private research projects. Based on these research findings, he has published several books and articles in journals of international impact. ORCID: https://orcid.org/0000-0002-6918-6910 Contact: rafael.pedraza@upf.edu

Lucía García-Carretero is a journalism graduate from the University of Valladolid and the holder of a Master’s in Social Communication from the Pompeu Fabra University (UPF). As the recipient of a research grant, she forms part of the Teaching and Research Staff at the UPF. She is a member of the Journalism Research Group and the Digital Documentation and Interactive Communication (DigiDoc) Research Group. Her research interests lie in the study of online political communication and the analysis of electoral communication strategies in digital networks. ORCID: https://orcid.org/0000-0002-1414-3921 Contact: lucia.garcia@upf.edu
Preface

Serie Editorial DigiDoc
This series reports the direct findings of a number of research projects. Indirect findings are typically published in refereed journal articles, but given their length, there is usually insufficient room for direct findings. After several years these results might appear in monograph form or simply lie forgotten in a drawer. Current trends in academic policy favour an open-access approach, whereby researchers are encouraged to make their results as widely available as possible, for example under Creative Commons licences, and where appropriate in institutional repositories or in the research group’s own repository. In keeping with this philosophy, we present this Serie Editorial and other forms of open-access dissemination that our group has adopted as part of the journal, Anuario Hipertext.net.

Active Audiences deliverables
The Active Audiences Project is concerned with the analysis of various aspects of the cyber media. The different activities that make up this project – “Active audiences and journalism. Interactivity, web integration and findability of journalistic information”, funded by the National R&D+i Plan, have generated results that have been published in indexed journals and presented at various conferences. However, they have also generated direct results. The dissemination of these direct results, in all cases related closely to our research objectives, is achieved via this collection of Deliverables, in keeping with open-access recommendations and guidelines regarding the need to make direct research results available too. This present deliverable corresponds to one of our secondary research lines, namely our focus on SEO and Communication.
A framework is a standardized set of concepts, practices, and criteria for dealing with a common type of problem, which can be used as a reference to help address and resolve new problems of a similar nature (Wikipedia).

0. Understanding this Framework

As the opening quotation above indicates, we understand that a framework should be able to provide both concepts and criteria. Here, in the case that concerns us, we would add that a framework should also provide a basic premise and a statement of its objectives and its overall scope.

The structure of the SEO-WCP Framework that we propose here is best summed up, therefore, in the following table:

<table>
<thead>
<tr>
<th>Nº</th>
<th>Component</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Objectives</td>
<td>To establish the objectives of the framework and its overall scope, based on the methodology adopted, as well as to outline the principal sources drawn upon.</td>
</tr>
<tr>
<td>2</td>
<td>Premise</td>
<td>To acknowledge a clear preference for the primacy of journalistic criteria over those of SEO in cases of incompatibility. To identify the different stages in the procedure. To provide a justification for the name given to the framework. To establishing a taxonomy of the differences between SEO and Journalism.</td>
</tr>
<tr>
<td>3</td>
<td>Concepts</td>
<td>To support and aid interpretation of the procedure with proposals for definitions of OnPage factors, including keywords and related concepts.</td>
</tr>
<tr>
<td>4</td>
<td>Procedures</td>
<td>The most characteristic features of the framework, namely, the recommended procedures and criteria.</td>
</tr>
</tbody>
</table>
1. Objectives, scope and sources

Our aim in the sections that follow is to propose a framework for optimizing journalistic content from two perspectives: that of search engine optimization (SEO), on the one hand, and that of visibility on social networks, on the other.

However, we ought to clarify from the outset that it is not our aim to provide a descriptive proposal, that is, we do not seek to describe how to optimize de facto the journalistic content of media sites in relation to SEO.

Indeed, we do not know exactly how they do it. A descriptive analysis of this kind would require, if it were to be attempted, a specific study of at least one media group and, even then, we do not know if the disclosure of this strategic information by the news media is possible. In any case, this is not something we seek to undertake at this stage in our work. However, whatever it is they do, what we do know is that it probably involves a variation of the framework that we present here.

We understand, therefore, that our framework is at the same time both possible and plausible as it is consistent and fully concurs with the tutorials and recommendations emanating from the best practices of so-called white hat SEO (including those of Google itself) as well with those originating from organizations linked to SEO as applied to the world of journalism, among which we should highlight the News University of the Poynter Institute (in this case via its webinars) and the BBC Academy, especially as regards such elements as the double headline system employed in this framework.

Our proposal is, therefore, independent of any specific news media site and, moreover, it is we believe compatible with the different writing styles of all the news media.

Our framework also draws on the two-stage optimization system, comprising writing + checking, developed by the producers of what is currently the most prestigious optimization software, YOAST SEO, one of whose extensions includes a plugin for optimizing news stories for Google News, which is also taken into consideration here.

We have also drawn on the reports of some of the leading SEO analysis companies, most notably those of Moz and Searchmetrics, as well as the results published in one of the chapters of the PhD thesis written by Dr. Carlos Gonzalo, supervised by one of the designers of the present framework. This study undertakes a systematic analysis of the more than 200 web positioning factors that Google uses and will form the basis of a forthcoming publication in the DigiDoc Series.

In addition, we have conducted an exhaustive review of recent publications on SEO factors as well as of the (few) academic publications that link SEO and Journalism, most notably the
contributions of Giomelakis and Veglis (2015, 2016). All these references can be found in the bibliography at the end of this paper.

Finally, the framework proposed here has benefited from a series of seminars and meetings involving SEO and Communication experts organized as part of the Active Audiences project. All other sources drawn upon are included in the bibliography.

At the outset we stressed that our proposed framework does NOT constitute a description of how the media, de facto, go about optimizing their content for search engines. We should now, therefore, identify the purposes we believe it can serve:

- A methodology for training and for the acquisition of skills in the field of SEO and Communication.
- A proposal for independent work, but one that is at the same time adaptable and compatible with the style books of different news media.
- A proposal for new cyber media that need to adopt a methodology to begin their professional activities.
- A useful proposal for comparing standards and procedures that any researcher or firm can adopt as part of their studies to improve their procedures.

2. Premise and naming of the framework

In this section we explain our working premise and the name chosen for the framework, the WCP Framework, given that the two concepts are related.

The initials that make up the name correspond to the three recommended phases of the optimization procedure:

[WRITE] > [CHECK] > [PUBLISH]

2.1. Premises

These three phases correspond to the principal premises of our framework, where they are related to the concepts of <Writing the news story> and <SEO check>. Thus we argue that that:

- the writing of the news story must adhere to the principles of journalistic style;
- the SEO check, as is logical, must adhere to criteria of search engine visibility and visibility on social networks;
- all possible contradictions must be resolved in favour of journalistic criteria (hence the primacy of journalistic principles);
once the news story has been written in accordance with the principles of journalism and the story has been optimized for search engines (and all contradictions have been resolved), it can be published or disseminated via various platforms or channels.

The following diagram illustrates the dual nature of these premises + phases, which we label with the initials WCP

![Diagram of the premises and the order of the phases in the WCP Framework](image)

**Figure 1: Diagram of the premises and the order of the phases in the WCP Framework**

### 2.2. Taxonomy

An initial taxonomy of the reasons for the differences between the journalistic principles and those of SEO principles points to two main elements, namely, **Context** and **Divergences**.

In keeping with these, the reasons for the need for different news headlines and SEO titles can be identified as follows. Thus, as regards the headline we need to recognise:

- **Context of the news story**:
  - As part of the cyber media web
  - As part of the search engine results page (SERP)
  - As part of a social network wall or timeline
- **Divergences**
  - Section and explicit headline vs Webpage context
  - Explicit components vs Webpage context
  - Surnames vs Full names
  - Style guide recommendations vs Search trends
The above taxonomy can be interpreted as follows: as far as the context is concerned, this is very different if we consider the news story in the context of its webpage, where it will appear under a clearly identified section, as opposed to in isolation as part of the wall of a Facebook page or a Twitter account timeline.

In turn, these differences of context give rise to the need to include the name of the section in which the news story appears within the SEO title (in the metadata <title>), something that is not necessary in the news site headline (<h1> tag); the same occurs with the need to use full names (name and surname) in the metadata (<title>), but not in the news site headline (<h1>), etc.

Having drawn these distinctions, in the following sections we present the different component parts, beginning with some conceptual elements.

3. Concepts

We divide the concepts in three parts: the double headline principle, platforms and keywords. We examine each of these below.

3.1. Double headline principle

This principle is closely linked to the premises (2.1) and to the taxonomy of reasons accounting for the divergences (2.2.). The principle informs us that thanks to the use of metadata the journalistic criteria and the SEO positioning criteria are compatible.

The double headline refers to the possibility/opportunity of using a news headline on the media site’s webpage with the <h1> tag in HTML – which fulfils the aforementioned journalistic criteria – and an SEO title modified in part using the metadata <title> – which fulfils the SEO criteria.

For example, a good working rule in cyber media is that the SEO title should always include the media brand name and the section title – two useful data for SEO, but unnecessary as part of the headline (h1) from the point of view of journalistic criteria.

This principle can be extended to the various distribution platforms of journalistic content, typically such digital networks as Facebook and Twitter.

3.2. Platforms

The webpage of a cyber media site is the natural place for the publication of its news stories. However, just because it is the “natural” place does not mean that the site’s webpage is the place where the news story will be seen for the first time nor where it will be read most frequently.
Indeed, we understand that today a news story or journalistic report is likely to be seen for the first time on at least one the following platforms, and which for the purposes of our framework can be considered the most important:

- Search Engine Results Page (SERP), where the news story will appear as a snippet in which the headline will be taken from the metadata &lt;title&gt; (not from the content of the h1 tag) and in which the description of the other metadata is also available.
- Facebook
- Twitter
- Webpage of a cyber media site

3.3. SEO factors and keywords

The different elements that have a positive or negative influence on the visibility and positioning of website, and therefore, of a news story, are referred to as the SEO factors. These factors may be fully controlled by the author of the page, when they are known as OnPage factors, or they may lie (partially) beyond the author’s control, when they are known as OffPage factors.

The main concern of OnPage optimization is the actual webpage content and here the keywords play a decisive role. Thus, in presenting our framework we begin by considering various aspects related to the vital concept of the keyword.

The point of view we adopt and the criteria on which we base our definition of keywords is journalistic, that is, the perspective corresponding to the newsroom or the journalist/author of the news story.

SEO factors - OnPage and OffPage

The features or properties of a page that help (or hinder) a site’s search engine positioning are known as SEO factors. These factors are divided into OnPage and OffPage factors. The former are those over which the creator of the page (in our case, the writer of the news story) has total control, as they are concerned with the page’s content. The latter lie (largely) beyond the control of the page’s creator, as they are concerned with inbound links received from other websites and from mentions on social networks by social actors. The framework presented here focuses exclusively on the OnPage factors.
**Keyword**

This is the term we hope will be used in searching for the news story or the small set of words which we hope will make the news story visible following a Google search. Normally, we optimize a story for **one** keyword. If we want to optimize the story for two or more words, then we must repeat the checks presented below and consider making the news story longer.

**Keyword density**

This is the number of times the keyword appears in relation to the total number of words contained in the news story. It is calculated by dividing the number of times the keyword appears by the total number of words and multiplying by a hundred. For example, if the keyword is used 10 times in a story that is 500 words long, then the density is 2 per cent.

**Optimum keyword density**

Is there an optimum density? Officially, Google says there is not. The official recommendation is that an article should be written for humans, not for Google, so that the best density is that which occurs spontaneously when using natural language. However, although there is no optimum density, a number of expert analyses clearly point to the existence of an optimum range as discussed below.

**Under-optimization**

Human beings are not always rational. We might write a long article under the impression that we have been expressing our ideas on a given topic, but in fact hardly use the keyword that best identifies that topic. In this case, the low keyword density will prevent Google from considering the news story relevant, leading to a case of search engine under-optimization, which is as unwelcome as over-optimization.
Over-optimization

Google’s analytic algorithms penalize over-optimization – that is, the unnatural repetition of the same word or phrase in a given text, above and beyond the usual frequency for a language. Good writing uses a combination of synonyms and expressly avoids repeating the same word too often in the same sentence. This natural quality of writing is what Google seeks to promote by penalizing texts over-optimized for search engines.

Optimum range

While it seems that there is no optimum density, there does appear to be an optimum range, which provides a clear indication of the minimum and maximum desirable densities. The minimum and maximum densities should be avoided so as not to run the risk of under-/over-optimization, respectively. According to various analyses, it seems that the optimum range is relatively wide, extending between 0.5 and 2.5 per cent, at least in the case of relatively long texts (300 words or more).

Optimum distribution

Many SEO professionals prefer to think in terms of an optimum distribution rather than of an optimum density. Hence, what is valued is where the keyword appears (i.e., in which parts of the page) and not how many times it actually appears. As long as the optimum range is respected, the idea of distribution is more efficient because it allows us to present a cogent entry structure, without abusing the natural language.

Latent Semantic Indexing

This expression originates from the theory of information retrieval, which search engines, in part, draw upon to “understand” the topic of a page. According to this theory, to know if a page matches with a certain keyword, the synonyms and related terms of that particular word must also be considered. For example, to know if a page matches a search that uses the term “human rights”, the search engine will deem pages to be more relevant if in addition to this term they also contain such words as “democracy”, “freedom”, “justice”, etc.
**Entities**

Individuals, organizations, cities, businesses, etc., constitute what are known as entities in semantic searches. More specifically, in the domain of the websites we refer of course to the actual names of entities. The appearance of such names and, therefore, the appearance of proper names, place names, etc., lends credibility to a news story, because it can be interpreted both as a reference to current newsworthy events as well as to possible sources.

Thus, entities constitute another way of addressing the concerns of the keywords that should appear in a news story.
4. Procedures

Let us stress from the outset that some of the above concepts can result in the need to edit or rewrite sections of the news story. These steps should always be taken, unless they affect the journalistic quality.

In the case of contradictions between the SEO imperative and the journalistic imperative, those of search engine optimization should always give way. Journalism is at the service of people, not of Google. For example, as we shall see below, checkpoint 1 proposes a minimum news story length of 300 words, but this should only be applied when it does not run counter to the specific journalistic genre.

That said, the points that the journalist/author should consider under this framework are:

01. **Length.** At least *300 words*, although it is much better if the news story exceeds this word count.

02. **Multimedia.** Always incorporate an element of multimedia, at least a photograph.

03. **Keyword.** Decide what the main keyword of the entry is.

04. **Optimum distribution** of the keyword

   i. In the main heading or the **news headline** (h1 tag)
   ii. In the **URL** of the entry. Edit the URL if necessary to avoid empty words and to limit length. Short, readily handled titles are preferable.
   iii. In the **SEO title**, via the <title> tag. Maximum 70 characters
   iv. In the **metadata** in the description of the <meta> tag. Maximum 156 characters
   v. In one of the headings in the body of the news story (h2 tag)
   vi. In the **first** paragraph
   vii. In one or more of the **central** paragraphs
   viii. In one or more of the **final** paragraphs
   ix. In one of the **images**
   x. In some of the external **hyperlinks** to related sources

05. **Emphasis.** In some of its occurrences, the keyword should be highlighted in bold or italics.

---

1One aspect that is sometimes not given the importance it should be given, and that can be influential, is the optimization of images with the appropriate use of tags; that is, adding a term to the image name in order that search engines can find it. The term should describe the image content and include a keyword so that it can be readily found by users. Additionally, the weight of the image needs to be taken into account, because if it is of high quality and heavy, the downloading speed of the webpage decreases, and search engines penalize sites that take too long to download their elements (Iglesias-García & Codina, 2016).
06. **Internal navigation.** If possible, the keyword should be linked to a category or internal tag.

07. **Semantic support.** Use synonyms of the keyword and related terms in the body and/or in the headings of the entry to strengthen the keyword.

08. **Credibility.** Mention entities in the form of the name of persons, places and institutions, and add links to the entities where appropriate.

09. **Internal links.** Wherever possible, establish links with other thematically related entries, using the website’s taxonomy or system of tags or categories.

10. **Social web and adaptive content.** Be sure to configure the entry so that it is published on the social networks, and that there are buttons of diffusion in activated and configured social networks. If the CMS so allows, consider titles, descriptions and specific images for networks such as Facebook and Twitter.

11. **Programming.** Program the entry, if appropriate, so that it is published at the optimum time in relation to its content and nature.

5. **Conclusions**

The framework presented here comprises a working combination of premises, concepts and procedures. It is not based on the analysis of how the processes of *de facto* news media optimization should be carried out, but rather on how it is plausible and at the same time feasible to consider the way in which these processes can be undertaken based on the best evidence available and on the best practices known.

The framework should be useful, not only as a tool for training and the acquisition of skills, but also as a starting point for new cyber media sites that need just such a spring board to optimize their journalistic content in terms of web visibility and SEO.

It should also serve as a starting point for comparative analyses and benchmarking studies for SEO and communication, since it offers a series of points that lend themselves to comparative study.

6. **Bibliography**


01.04.2015.


IGLESIAS-GARCÍA, Mar; CODINA, Lluís (2016). "Los cibermedios y la importancia estratégica del posicionamiento en buscadores (SEO)". **Opción**. In press.


MORATO, Jorge; SÁNCHEZ-CUADRADO, Sonia, MORENO-PELAYO, VALENTÍN, & MOREIRO, José Antonio. 2013. "Evolución de los factores de posicionamiento web y adaptación de las herramientas de optimización". **Revista española de documentación científica**. Vol.36. Nº3. Available at: [http://dx.doi.org/10.3989/redc.2013.3.956](http://dx.doi.org/10.3989/redc.2013.3.956) Consulted on 05.06.2015


