

---

# Técnicas digitales para el estudio de las competencias y perfiles profesionales: el caso de la oferta laboral de SEO

*Digital techniques for the study of professional skills and profiles: the case of SEO job offers*

---

**Raquel ESCANDELL-POVEDA, Natalia PAPI-GÁLVEZ, Mar IGLESIAS-GARCÍA**

Universidad de Alicante, Carretera San Vicente del Raspeig s/n, 03690 San Vicente del Raspeig, Alicante, España, r.escandell@ua.es, natalia.p@ua.es, mar.iglesias@ua.es

## Resumen

La relevancia de los medios digitales requiere de nuevas metodologías para su investigación dada la naturaleza del entorno online. En la cultura digital, el conocimiento del mercado laboral tiene implicaciones tanto para la economía como para la educación por su compromiso de fomentar la empleabilidad entre el estudiantado. En el contexto de la comunicación digital, surgen nuevos perfiles que pueden ser estudiados a partir de los requerimientos que empresas e instituciones demandan a partir de las ofertas de empleo. Se diseña un marco metodológico orientado a investigaciones que requieran el análisis de la demanda laboral, a partir la recolección, extracción y volcado del contenido de las ofertas de trabajo, asociadas a una de las nuevas figuras profesionales surgidas en comunicación digital: el SEO o especialista en posicionamiento en buscadores. Se expone el método que combina fases manuales y automatizadas así como el empleo de técnicas digitales como el *web scraping* para la descarga masiva de datos. Gracias a esta técnica, se convierte la información desestructurada proveniente de LinkedIn y el portal de empleo Infojobs en un corpus de datos ordenados a partir del cual realizar análisis cuantitativos o cualitativos. En este artículo, se evidencia la utilidad del método y su aplicabilidad en futuros estudios sobre características, competencias y requerimientos laborales. Igualmente se detectan limitaciones que plantean la urgencia de mejorar las herramientas en el marco de la investigación social.

**Palabras clave:** Web scraping. Perfiles profesionales. Metodologías digitales. SEO. Mercado laboral. Recogida de datos. Acceso a la información.

## 1. Introducción

El método es clave para la generación de conocimiento científico. La aceptación de los resultados de un estudio por parte de la comunidad científica dependerá, en primera instancia, de la aplicación adecuada de los métodos, definidos según el campo de conocimiento al que pretende contribuir. La investigación social, en particular, provee de un conjunto amplio de técnicas que han sido suficientemente debatidas y contrasta-

## Abstract

The relevance of digital media requires new methodologies for research given the nature of the online environment. In the digital culture, knowledge of the labor market has implications for both the economy and education because of its commitment to promote employability among students. In the context of digital communication, new profiles emerge that can be studied from the requirements that companies and institutions demand from job offers. A methodological framework is designed for research that requires the analysis of labor demand, based on the collection, extraction and downloading of the content of job offers, associated with one of the new professional figures that have emerged in digital communication: the SEO or search engine optimizer. The method that combines manual and automated phases is presented, as well as the use of digital techniques such as web scraping for massive data downloading. Thanks to this technique, unstructured information from LinkedIn and the Infojobs employment portal becomes a corpus of ordered data from which to perform quantitative or qualitative analyses. In this paper, the method evidences its usefulness and its applicability in future studies on characteristics, competencies and labor requirements. Limitations are also detected that raise the urgency of improving the tools within the framework of social research.

**Keywords:** Web scraping. Professional profiles. Digital methodologies. SEO. Labor market. Data acquisition. Access to information.

das durante años, desde su paradigma cuantitativo o cualitativo, para abordar los fenómenos sociales que persigue explorar, describir, explicar, confirmar o evaluar. La utilización de tales técnicas queda inexorablemente sujeta al cumplimiento de ciertos requisitos en su aplicación, suficientemente conocidos entre los científicos sociales, como también lo son sus fortalezas y sus limitaciones.

A este respecto, la cantidad y diversidad de fuentes, el acceso a las mismas y el procesamiento

de la información son elementos cruciales para comprender las posibilidades que tiene la investigación social en su empeño por abordar los diferentes acontecimientos, con especial atención en los entornos digitales y, en particular, los vinculados al campo de la comunicación.

La sociedad digital despierta la necesidad de estudios que ayuden a comprender los cambios que se están produciendo en el ecosistema mediático y en sus industrias. Ante este escenario, las técnicas no solo se adaptan a los canales digitales, también precipitan nuevas soluciones. La relativa facilidad de acceso a la información de internet y la operacionalización de los procesos, multiplican las fuentes y aceleran las primeras etapas del estudio, tales como la captura y el registro de los datos.

La metodología digital aplicada al campo de la comunicación se ubica en este lugar, alcanzando su máxima expresión en la defensa del surgimiento de una ciencia computacional de la comunicación que “está brindando más oportunidades a los investigadores para abordar preguntas de investigación nuevas y antiguas, y esto, a su vez, contribuirá a un cambio en la calidad de las investigaciones en medios y comunicación” (Arcila-Calderón, et al., 2021, p. iv). Con todo, en su esencia, la metodología sigue respondiendo a una forma de organización de un conjunto de elementos “para convertirlos en información (saberes) sobre el funcionamiento del mundo” (Flores-Márquez y González-Reyes, 2021, p. 16). Así, las técnicas que quedan incluidas actualmente bajo el paraguas de esta metodología aplican procesos automatizados, en algunos casos llevados a cabo con ayuda de herramientas específicas, pero no sustituyen, aún, a la figura del investigador.

Esta consideración se sustenta en el hecho ineludible de que el objeto de la investigación trasciende la mera captura y gestión de los datos. A este respecto, y especialmente en los últimos años, se ha escrito profusamente sobre la gestión de los datos masivos, aludiendo a diferentes Vs del *Big Data* como punto de partida para comprender las características de la producción de datos de este nuevo entorno, y que se fueron sumando a las tres iniciales (volumen, variedad y velocidad) de Doug Laney (2001). Sin embargo, también debe incorporarse el “valor” como rasgo definitorio del *Smart Data*, con el propósito de que los datos se transformen en información y que la información derive en conocimiento (Lafate, 2014; Valle y De la Pena, 2014; Leonelli, 2020), necesario para el avance de cualquier ciencia. En este proceso, es necesaria la preparación de las personas que tienen la competencia investigadora para que testeen nuevas fórmulas

en entornos cuyas fuentes de información no necesariamente son académicas aplicando, en su utilización, los preceptos científicos aprendidos. Esta cuestión se hace especialmente patente cuando se investiga en publicidad, por su proximidad al sector profesional.

### 1.1. Técnicas digitales, *web scraping* y *crawling*

Hay una gran variedad de técnicas digitales para la investigación social, que han sido clasificadas y catalogadas desde diversas perspectivas (Rodríguez-Cano, 2022). Entre estas técnicas destaca el *scraping* (raspado o extracción de datos), una de las vertientes de la minería de datos, que permite la recolección automática de información en internet. Mediante la ejecución de un software específico, y a partir del objetivo de la investigación, se pueden recuperar aquellos datos que se han definido con anterioridad (Rodríguez-Cano, 2022).

En el caso del *web scraping*, (también denominado *screen scraping* o *information scraping*), se trata de un proceso automatizado que usa *bots*, *spiders* o *crawlers*, para acceder a documentos en internet, buscar la información predefinida, recopilarla y después descargarla, transformada en un formato estructurado (Massimino, 2016). Por otro lado, el *web crawling* o rastreo web es la navegación sistemática y automatizada de la que se extrae la URL, pero no la información específica. El contenido completo está disponible a través del hipervínculo, pero generalmente no se archiva (Hillen, 2019). Estas técnicas han sido empleadas en múltiples estudios de investigación social (Marres y Weltevrede, 2013; Pereira y Vanitha, 2015; Dewi y Chandra, 2019; Milovanović et al., 2022; Morales-i-Gras y Sánchez-i-Vallès, 2022; Wang, 2022; Wiechetek y Pastuszak, 2022; Zarrabeitia-Bilbao, 2022), evidenciando su utilidad como parte de la metodología imprescindible en este entorno digital donde el contenido está en perpetua actualización.

Las técnicas digitales deben distinguirse de las denominadas herramientas digitales: el software utilizado para aplicarlas. Las técnicas son procedimientos validados que operan en un nivel conceptual, mientras que las herramientas son programas informáticos. Habitualmente hay más de una herramienta para una técnica y la elección de una u otra dependerá de las necesidades de la investigación y sus parámetros (Sued, 2020).

Las herramientas creadas para hacer *web scraping*, para rastrear, recoger y estructurar la información requerida son numerosas y de diferente tipo: herramientas para programación, aplicaciones web, instaladores y extensiones para nave-

gadores. Ejemplos de herramientas de programación son *BeautifulSoup* y *Request*, así como *Scrapy*. Para usar estas herramientas es necesario tener conocimientos de programación. Más sencillas de usar son las aplicaciones de escritorio, entre las que destacan *Octoparse* y *ParseHub*, mientras que las extensiones para navegadores (plugins) son las más rápidas (Moo-Herrera et al, 2020). Entre estas últimas se encuentran *Link Klipper* y *Web Scraper*. Por un lado, *Link Klipper* es una extensión que se instala en el navegador y que permite descargar todos los enlaces de una página web. Es un complemento público y gratuito disponible para Google Chrome que permite extraer los enlaces y exportarlos como un archivo CSV. *Web Scraper* también se utiliza a través de su extensión para Google Chrome, disponible asimismo para *Firefox*, y permite descargar datos que están en la web.

## 1.2. Cultura digital y mercado laboral: el SEO

La cibercultura o cultura digital engloba a “las culturas que están emergiendo a raíz del uso generalizado y ubicuo de las TIC, donde el ocio, la comunicación, la política, la educación o la economía se encuentran fuertemente mediadas tecnológicamente” (Lasen y Puente, 2016, p. 6). En la parcela económica podríamos enmarcar el mercado laboral, donde la permanente conectividad digital toma protagonismo en el modo de relacionarse con el trabajo, especialmente en sectores integrados en la economía digital (Lasen y Puente, 2016). Esta evolución trasciende al proceso de selección y búsqueda de empleo. Muestra de ello es que en España casi el 60 % de las compañías consultan las redes sociales de un futuro candidato, fundamentalmente su perfil en la red social LinkedIn (Infojobs, 2022). Asimismo, uno de cada seis empleados tiene una cuenta profesional en redes sociales y más de la mitad de ellos la utilizan para estar al día de las oportunidades laborales y noticias del mercado (Infojobs, 2022). Diversas investigaciones académicas han utilizado tanto estas redes sociales como los portales de empleo para la detección de competencias profesionales en perfiles relacionados con la comunicación y las nuevas tecnologías (Morato et. al, 2016; Álvarez-Flores et. al, 2018; Jiménez-Gómez y Mañas-Viniegra, 2018; Clemente-Mediavilla y Antolín-Prieto, 2019; Marta-Lazo et. al, 2018). Las pretensiones de estos estudios van encaminadas a detectar las necesidades reales de las empresas y los requisitos solicitados a sus candidatas para entrar en ellas. Sus resultados tienen implicaciones directas en otro de los ámbitos de la cultura digital, la enseñanza, pues, tras la entrada en el Espacio Europeo de Educación Superior, las instituciones académi-

cas, junto con las empresas, adquieren el compromiso de fomentar la empleabilidad de los egresados y garantizar la adquisición de competencias profesionales que mejoren su inserción laboral (Ministerio de Asuntos Económicos y Transformación Digital, 2021).

El aumento de penetración y consumo de canales digitales da lugar a la aparición de nuevos perfiles profesionales especializados en comunicación y marketing digital, que responden a las necesidades de empresas e instituciones de alcanzar a sus audiencias a través del canal online. En este contexto surge el perfil de especialista en posicionamiento web o SEO, por sus siglas en inglés, *Search Engine Optimizer*, persona encargada de atraer a esos targets a través de los buscadores. El SEO es considerado uno de los perfiles en auge en el sector de la comunicación y marketing digital (The Social Media Family, 2017) y sus múltiples referencias en estudios relativos a competencias y nuevas figuras profesionales determinan su relevancia (Boettger et al., 2017; Saura et al., 2017; Miguel-San-Emeterio, 2018; Popescu et al., 2019; Giomelakis et al., 2019; Lopezosa et al., 2020; Sheffield, 2020). Se considera pertinente, por lo tanto, estudiar las características de esta figura profesional, acudiendo para ello a una fuente pública y accesible que refleja las necesidades del tejido empresarial y permite el estudio, no solo de competencias, sino de otros requerimientos relacionados con este perfil: las ofertas laborales.

## 2. Objetivos

El principal objetivo de este artículo es diseñar un marco metodológico aplicable a investigaciones que requieran el análisis de la demanda laboral a partir del contenido de ofertas de trabajo publicadas en plataformas digitales.

Como objetivos secundarios se encuentran los de exponer los procesos y técnicas empleadas para la recolección, extracción y volcado del contenido de estas ofertas laborales alrededor de una competencia o figura profesional. Por un lado, se describen las fases en las que se lleva a cabo el método y, por otro, se indican las herramientas empleadas y se detalla su aplicación práctica para la consecución de los resultados. Concretamente se pretende mostrar el funcionamiento de las técnicas de *web scraping* y automatización de procesos en la extracción de datos para la investigación de perfiles profesionales y evaluar su utilidad. Asimismo, trata de ejemplificar el método incluyendo las etapas y especificaciones concretas para su aplicación a partir de las dos fuentes de ofertas de empleo más importantes en España: LinkedIn e Infojobs.

El producto resultante de la aplicación de este método es un corpus de datos, ordenado y estructurado que sirve de punto de partida para realizar estudios cuantitativos o cualitativos de las características, competencias y requerimientos solicitados por parte del tejido productivo, a través de sus ofertas de empleo en un determinado periodo de tiempo.

### 3. Metodología

Para la consecución de los objetivos propuestos, se ha llevado a cabo el desarrollo empírico del método en un caso concreto: la demanda laboral del SEO en España. Se ha elegido el SEO por ser uno de los perfiles más relevantes en el escenario de la comunicación digital y por su versatilidad, al facilitar la demostración de este método cuando se persigue la extracción tanto de perfiles profesionales, solicitados en las ofertas, como la identificación de las competencias. El SEO puede considerarse tanto un perfil, como una competencia asociada a otras figuras profesionales, en cuyo caso las siglas hacen referencia a *Search Engine Optimization*. La delimitación de esta competencia o perfil se realiza a partir de la estrategia de búsqueda inicial, lo que permite una sencilla extrapolación a otras figuras o requerimientos competenciales.

Las fuentes elegidas para realizar la investigación han sido LinkedIn e Infojobs. Según el Informe Talento Conectado publicado por Infoem-

pleo y EY (2018), LinkedIn es la principal red social en España de búsqueda de empleo. Infojobs, por su parte, es el portal de empleo más importante del país, con 324.448 ofertas publicadas en 2017 (TMP Worldwide), solo por detrás de LinkedIn, en la que aparecieron 589.577, el 44% del total de las publicadas en el territorio nacional.

Tras la selección del objeto de estudio y las fuentes se llevaron a cabo diferentes técnicas manuales y automáticas y se evaluaron diversas herramientas para la consecución de los resultados hasta dar con las más óptimas, a partir de las cuales se expone este marco metodológico. Son consideradas las decisiones que afectan al silencio o al ruido de la estrategia de búsqueda en tanto conceptos fundamentales para la evaluación de cualquier procedimiento de recuperación (Van Rijsbergen, 1979). El periodo en el que se llevó a cabo el trabajo de campo de recuperación de ofertas comprendió seis meses completos, desde febrero a julio de 2019, ambos incluidos.

### 4. Resultados

El proceso metodológico empleado se conforma de una serie de fases que combinan una parte manual y otra automática, derivada de las posibilidades tecnológicas que permite el medio. En la Figura 1 se encuentra sintetizado, a modo de diagrama que integra las dos fases principales: selección y descarga de ofertas y publicación de páginas y *scraping* de contenidos.

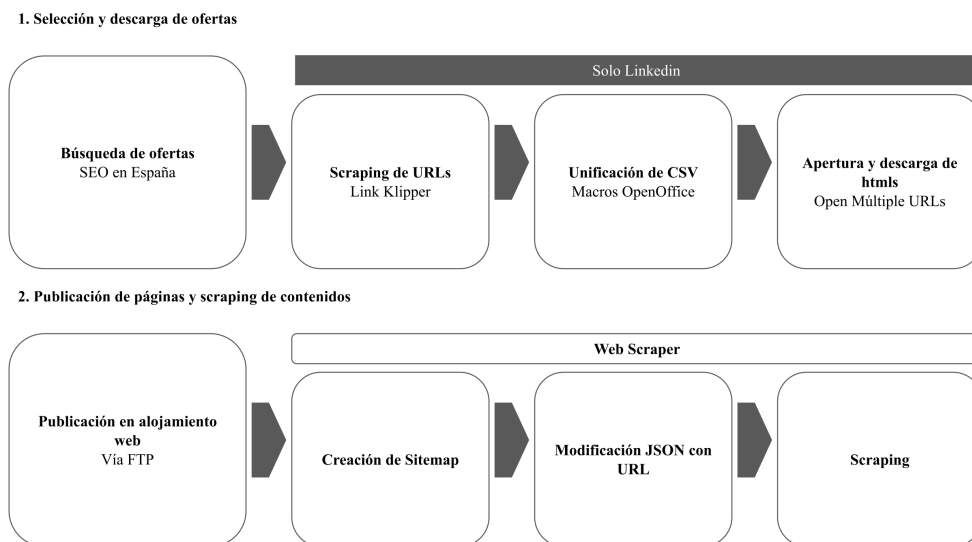


Figura 1. Procedimiento de descarga y scraping de ofertas laborales

#### 4.1. Selección y descarga de las ofertas

La primera fase consiste en establecer la estrategia de búsqueda de las ofertas que se repite mensualmente a lo largo del periodo estudiado.

Esta depende, en gran medida, de la plataforma en la que nos encontremos. En el caso de la red social LinkedIn, el buscador de su sección laboral

permite realizar una búsqueda libre y además filtrar según los parámetros de fecha de publicación de la oferta, nivel de experiencia, nombre de la empresa, tipo de trabajo, localización o industria, entre otros. En lo que respecta al portal de empleo Infojobs, su buscador principal dispone de una casilla de texto libre y la opción de elegir la ubicación. Dado el objeto de la investigación, la estrategia se limitó a buscar el término “SEO” y la ubicación de España en ambas webs. Sobre la restricción temporal, en el caso de LinkedIn, a partir del segundo mes se filtró la búsqueda a las ofertas publicadas “en el último mes”, con el propósito de evitar acumular las ya recuperadas. En Infojobs, a partir de la segunda búsqueda, se tuvieron en cuenta las ofertas etiquetadas como “nuevas”, con el mismo propósito.

El siguiente paso consiste en la descarga y almacenamiento de las ofertas. Sin embargo, no se puede hacer rápidamente porque los portales seleccionados no ofrecen una base de datos preparada para su análisis social. Además, solo en la red social LinkedIn, la estrategia de búsqueda empleada arrojaba un gran volumen de ofertas, alrededor de 700 mensuales; por lo que se planteó la automatización en la recuperación de la información en esta etapa. La automatización se considera justificada, además de por el volumen de información, por la naturaleza efímera de la misma, ya sea por su caducidad en el propio portal o por su eliminación por parte de la empresa ofertante. Sin embargo, aquí también hay limitaciones, máxime si se pretende aplicar el *scraping* de contenidos para poder analizar varios campos; pues los portales pueden tener restricciones de acceso si detectan procesos automáticos en la navegación por su sitio. En consecuencia, se hace precisa la preparación previa de la información a través de la aplicación de varias soluciones, según la presentación de la información del portal.

En la red social LinkedIn, se utiliza una herramienta de *scraping* para la descarga de todos los enlaces a las ofertas que aparecen en los resultados de búsqueda. Concretamente, se emplea la extensión para el navegador *Google Chrome* denominada *Link Klipper*. Tras visitar varias de las páginas de las ofertas resultantes, se detectó que todas ellas contenían en su URL la cadena de caracteres: “*NotAvailableFromVoyagerAPI*” no incluida en otros enlaces externos e internos de la web. Esta característica ayuda a mejorar la precisión en el proceso de *scraping* de direcciones, ya que la herramienta permite filtrar las URL a descargar aplicando expresiones regulares. Para ello, se accede a sus opciones de configuración (*Settings*) y se incluye el fragmento en su opción de filtro por *Regex*. Desde este mismo panel de configuración se puede elegir descargar

los archivos como extensión .csv para trabajarlos posteriormente en un programa de hojas de cálculo (Figura 2).

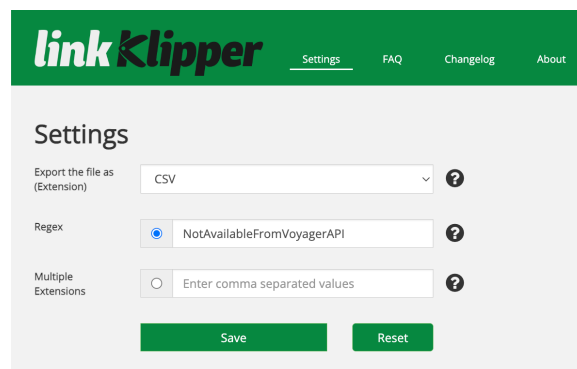


Figura 2. Panel de configuración de la herramienta *Link Klipper* (*Link Klipper*).

Una vez establecidos los parámetros de configuración, se procede a la descarga de enlaces desde las páginas de resultados. Se recomienda hacer *scroll* en cada una de ellas para asegurarse de que se ha cargado por completo. Una vez al final de la página, se abre la extensión y se pulsa: *Extract All Links*, para la descarga del archivo csv con las URL de las ofertas. Este proceso se repite en todas las páginas resultantes tras la búsqueda de ofertas de empleo que contienen SEO.

```

1 Sub Main
2 Dim options(0) As New com.sun.star.beans.NamedValue
3
4 util = createUnoService("org.universolibre.EasyDev")
5 source = createUnoStruct("org.universolibre.EasyDev.CellRangeAddress")
6 target = createUnoStruct("org.universolibre.EasyDev.CellRangeAddress")
7
8 'Obtenemos todos los archivos ODS de la ruta especificada
9 files = util.GetFiles("/home/mau/Downloads/Docs", "ods")
10 'Generamos un nuevo documento destino
11 new_doc = util.newDoc("")
12 target.Doc = new_doc
13 'Abrimos los documentos de forma oculta
14 options(0).Name = "Hidden"
15 options(0).Value = True
16 For i = LBound(files) To UBound(files)
17     'Abrimos el documento
18     doc = util.openDoc(files(i), options)
19     source.Doc = doc
20
21     'Copiamos todas las hojas del origen al destino, si tienes OpenOffice
22     'util.sheetCopyToDoc(source, target, -1, True, False)
23     'Copiamos todas las hojas del origen al destino, si tienes LibreOffice
24     util.sheetInsertFromDoc(source, target, -1, False)
25     doc.dispose()
26 Next
27 End Sub

```

Figura 3. Código del macro para unificar las páginas en *OpenOffice Calc* (Baeza-Servin, 2015, diciembre 3)

Cada página de resultados se descarga como un archivo diferente, por ello, el siguiente paso es la unificación de todos los enlaces de ofertas en una misma tabla, previa al vaciado automatizado de la información. Para ello, aplicamos macros en el programa de gestión de hojas de cálculo de código abierto *OpenOffice Calc*. Las macros son secuencias desarrolladas en códigos de programación que permiten automatizar tareas (Blog Open-Office, 2015). En esta ocasión, utilizamos

una macro para combinar todos los archivos en uno y otra para fusionar todas las pestañas de ese archivo en una sola tabla. Primero instalamos la extensión *EasyDev* al programa, que simplifica la programación de macros, y luego creamos nuevas desde la opción *Herramientas > Macros > Organizar Macros*, a partir del código de Baeza-Servin (2015) para combinar los archivos y Federico (2017) para fusionar las pestañas (Figuras 3 y 4).

```

1 Sub UnirHojas()
2 Dim document as object
3 Dim dispatcher as object
4 Dim args1(5) as new com.sun.star.beans.PropertyValue
5 Dim args3(0) as new com.sun.star.beans.PropertyValue
6
7 document = ThisComponent.CurrentController.Frame
8 dispatcher = createUnoService("com.sun.star.frame.DispatchHelper")
9 args1(0).Name = "sel"
10 args1(0).Value = true
11
12 ThisComponent.Sheets.InsertNewByName("UnionDeDatos",0)
13 ThisComponent.CurrentController.Select(ThisComponent.Sheets(0).GetCellRangeByName("A1"))
14
15 For a = 1 to ThisComponent.Sheets.GetCount()-1
16 ThisComponent.CurrentController.Select(ThisComponent.Sheets(a).GetCellRangeByName("A1"))
17 dispatcher.executeDispatch(document, ".uno:GoToEndOfData", "", 0, args1())
18 dispatcher.executeDispatch(document, ".uno:Copy", "", 0, Array())
19 args1(0).Name = "Nr"
20 args1(0).Value = 1
21 dispatcher.executeDispatch(document, ".uno:JumpToTable", "", 0, args1())
22 args1(0).Name = "Flags"
23 args1(0).Value = "sv"
24 dispatcher.executeDispatch(document, ".uno:InsertContents", "", 0, args1())
25 dispatcher.executeDispatch(document, ".uno:GoToEndOfData", "", 0, args1())
26 dispatcher.executeDispatch(document, ".uno:GoDown", "", 0, args1())
27 dispatcher.executeDispatch(document, ".uno:GoLeftToStartOfData", "", 0, args1())
28 Next
29
30 ThisComponent.CurrentController.Select(ThisComponent.Sheets(0).GetCellRangeByName("A1"))
31 dispatcher.executeDispatch(document, ".uno:GoToEndOfData", "", 0, args3())
32 args1(0).Name = "aExtraWidth"
33 args1(0).Value = 100
34 dispatcher.executeDispatch(document, ".uno:SetOptimalColumnWidth", "", 0, args1())
35 End Sub

```

Figura 4. Código del macro para fusionar pestañas en OpenOffice Calc (Federico, 2017, enero 27)

Una vez que se tienen todas las URL listadas en la misma tabla, se deben eliminar las duplicadas.

Finalmente, con el listado ordenado y sin duplicados, se procede al último paso de esta fase: la apertura de las páginas y su descarga en el ordenador. Para agilizar el proceso, se puede utilizar una extensión del navegador *Google Chrome* denominada: *Open Multiple URL*, que permite abrir de manera simultánea un listado de URL en diferentes pestañas. Tras su apertura, se entra en cada una de las pestañas y se guarda la página como html desde la opción integrada en *Google Chrome*. Para realizar este paso, hay que tener en cuenta que LinkedIn establece un límite de visualización de sus páginas, especialmente si se realizan con usuario registrado. Para solventar este inconveniente se debe dejar un intervalo de tiempo para continuar con las visualizaciones. En el momento de realizar este estudio, la cadencia temporal entre descargas múltiples era de 30 minutos.

Paralelamente, en lo que respecta al portal de empleo Infojobs, la cantidad de las ofertas es mucho menor que en LinkedIn, por lo que su apertura y descarga como archivo html se realiza directamente desde las páginas de resultados de búsqueda, sin el empleo de las extensiones ni las macros.

Este proceso se realiza mensualmente durante el periodo de trabajo de campo establecido, que en esta investigación se prolongó durante medio año. Una vez finalizado, la cantidad total de ofertas descargadas ascendió a 6.044: 4.667 en LinkedIn y 1.377 en Infojobs, cuyo desglose mensual se puede observar en la Tabla I. En LinkedIn, el reparto de la cantidad de ofertas en los primeros meses es orientativo debido a los obstáculos encontrados en la plataforma descritos y solucionados según se ha descrito.

En Infojobs, la cantidad de ofertas del primer mes es mayor, pues se descargan todas las que responden a la estrategia de búsqueda vigentes ese momento. En los meses siguientes, se restringe a las nuevas ofertas publicadas.

	LinkedIn	Infojobs
Febrero	762	431
Marzo	741	254
Abril	1.138	179
Mayo	693	212
Junio	556	161
Julio	886	140
N (total)	4.776	1.377
Errores en el proceso de descarga u ofertas inexistentes	109	0
N (válido)	4.667	1.377

Tabla I. Desglose mensual de ofertas recuperadas

#### 4.2. Extracción y volcado de la información de las ofertas con *web scraping*

Una vez se dispone de todas las páginas de ofertas, la siguiente fase consiste en volcar todo su contenido en una hoja de cálculo mediante la técnica de *web scraping* para facilitar su posterior análisis.

Como primer paso, se renombran todos los archivos html o htm de las páginas descargadas numerándolas, siempre diferenciando entre ambos portales, con nombres como 0001linkedin y 0001infojobs. El objetivo es facilitar su inclusión en el posterior proceso automatizado de extracción de contenidos. En el sistema operativo *MacOs*, numerar los archivos de una carpeta se puede hacer de manera automática seleccionándolos y entrando a la opción de "Renombrar" en el menú contextual.

A continuación, se publican todas las páginas de las ofertas en un alojamiento web propio mediante un programa FTP, que permite la transfe-

rencia de archivos entre el ordenador y el servidor. De este modo se posibilita su acceso a través del navegador, donde se encuentra la extensión de *web scraping* que se utilizará para la extracción de datos. Un ejemplo de URL de una página de oferta sería el nombre de dominio con su extensión, asociado al alojamiento web, seguido de una barra y el nombre de archivo: nombrededominio.com/0001linkedin.html.

Seguidamente se procede a la configuración de *Web Scraper*: la herramienta que va a permitir la extracción de datos mediante esta técnica de *scraping* o “raspado”.

*Web Scraper*, de *webscraper.io*, se utiliza a través de su extensión para el navegador, *Google Chrome*, también disponible para *Firefox*. Se encuentra integrada dentro de la consola de herramientas para desarrolladores por lo que, para acceder a ella es necesario entrar en la opción del menú contextual del navegador: *Más Herramientas*, y allí *Herramientas para Desarrolladores*. Una vez abierta, se ancla el panel en la parte inferior y desde ahí se selecciona la opción *Web Scraper*.

El primer paso en el proceso de *scraping* con esta herramienta es crear un mapa del sitio, *sitemap*, a partir del cual se realizará el raspado de datos de la página. Para ello se abre la opción de *Create new sitemap* y a continuación *Create sitemap*.

Hay que especificar un nombre, con el que identificaremos el mapa del sitio y la URL de la página sobre la que se va a realizar el *scraping*, como la dirección en la que se encuentra una de las ofertas: [www.nombrededominio.com/0001linkedin.html](http://www.nombrededominio.com/0001linkedin.html). Más adelante se añadirá el total de las URL, aunque también existe la opción de especificar un rango de URL si las tenemos numeradas, sustituyendo el número que aparece en la dirección por el rango de números delimitados por corchetes al que debe hacer referencia. Así: [www.nombrededominio.com/\[0001-4667\]](http://www.nombrededominio.com/[0001-4667]).

Una vez creado el *sitemap*, se deben añadir los selectores, es decir, los diferentes campos de los cuales extraemos la información. Como la configuración de la página de ofertas es diferente para cada portal analizado, se realizan dos *sitemaps* diferentes, con sus selectores correspondientes.

La selección de estos campos depende de los elementos que contenga la oferta de los que se quiera analizar la información. En el caso de las ofertas de LinkedIn e Infojobs, los datos que se analizaron de cada fuente fueron distintos dadas sus propias diferencias, siendo Infojobs la opción mejor estructurada y con más campos (al tratarse de una web específica para ofertantes y demandantes de empleo) y LinkedIn la que contaba con

un corpus de ofertas más amplio. Los elementos analizados en cada uno de los portales se muestran en la Tabla II.

	LinkedIn	Infojobs
Cantidad total de ofertas (que contienen SEO)	Sí	Sí
Ofertas por provincias	Sí	Sí
Ofertas exclusivas de SEO (que no contienen otros perfiles en el título)	Sí	Sí
Otros Perfiles incluidos en el título (adicionales a SEO)	Sí	Sí
Nivel del puesto	Sí	Sí
Nivel de estudios		Sí
Tipo de estudios		Sí
Número de personas al cargo		Sí
Tipología de contrato		Sí
Tipo de jornada	Sí	Sí
Salario		Sí
Sectores relacionados con el puesto	Sí	Sí
Sector al que pertenece la empresa	Sí	
Tamaño de la empresa (según número de empleados)	Sí	
Funciones del puesto de trabajo	Sí	
Conocimientos específicos requeridos		Sí
Competencias generales requeridas		Sí

Tabla II. Elementos analizados en las ofertas de LinkedIn e Infojobs

Se abre el mapa del sitio creado en la aplicación de *Web Scraper*. Una vez en él, se añaden los selectores (pulsando *Add new selector*, e incluyendo un ID, identificador único que será la cabecera de nuestra columna en la hoja de cálculo) y se especifica qué tipo de selector se quiere que, en este caso, fue texto. A continuación, con el puntero del ratón se pulsa sobre el ítem de la página que queramos capturar, por ejemplo, el título de la oferta, que nos permitirá filtrar posteriormente el número de ofertas exclusivas de SEO y los perfiles incluidos en las que no lo son. Una vez se pulsa sobre él, aparece su etiqueta html, en este ejemplo: *h1*, se hace clic sobre *Done Selecting* para finalizar y esta etiqueta pasa a la casilla de *Selector* (Figura 5, en la página siguiente).

Para comprobar que la información del *h1* es exactamente la seleccionada (en este caso, título de la oferta) se pulsa en *Data Preview*. Una vez comprobado, se guarda con clic en *Save Selector* y se realiza el mismo proceso con todos los elementos de la página de los que se quiere extraer información. En esta investigación, los campos



extraídos de las ofertas de LinkedIn fueron: título, empresa, ciudad, fecha de la publicación, descripción, nivel de antigüedad, tipo de empleo, funciones laborales (varios elementos) y sectores (varios elementos).

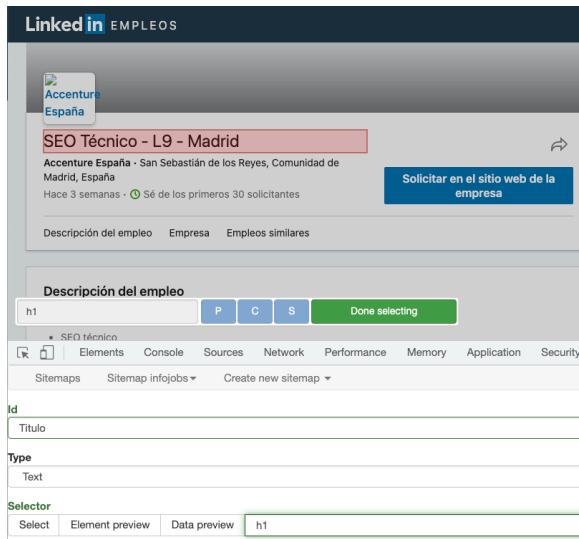


Figura 5. Captura del elemento título de una de las ofertas de trabajo en la herramienta Web Scraper (LinkedIn y Web Scraper)

Con respecto a las páginas de ofertas de Infojobs, el proceso es el mismo, aunque desde su propio *sitemap*. Los elementos configurados fueron: título, empresa, ubicación, fecha de publicación, descripción, nivel de experiencia, sector, tipo de contrato, tipo de jornada, salario, experiencia requerida, nivel de estudios requeridos, conocimientos (varios elementos), requisitos mínimos, personas al cargo, número de vacantes y horario. Los selectores configurados se van añadiendo al *sitemap* y pueden previsualizarse, editarse o eliminarse (Figura 6).

ID	Selector	type	Multiple	Parent selectors	Actions
Título	h1.text-hyphen	SelectorText	no	_root	<a href="#">Element preview</a>
Empresa	a.link.bold, div.col-6:nth-of-type(1) div.col-child > ul.list-default > li:nth-of-type(1)	SelectorText	no	_root	<a href="#">Element preview</a>
ubicacion	div.col-6:nth-of-type(1) div.col-child > ul.list-default > li:nth-of-type(1)	SelectorText	no	_root	<a href="#">Element preview</a>
fecha-publicacion	li:nth-of-type(3) span:nth-of-type(2)	SelectorText	no	_root	<a href="#">Element preview</a>
descripcion	div.inner.panel-canvas	SelectorText	no	_root	<a href="#">Element preview</a>
nivel-experiencia	span#prefijoNivelLaboral.list-default-text	SelectorText	no	_root	<a href="#">Element preview</a>
sector-tipo-industria	div.inner.panel-canvas li:nth-of-type(1) span.list-default-text	SelectorText	no	_root	<a href="#">Element preview</a>
tipo-empleo-contrato	span#prefijoJornada	SelectorText	no	_root	<a href="#">Element preview</a>
funciones-categoria	div.inner.panel-canvas li:nth-of-type(2) span.list-default-text	SelectorText	no	_root	<a href="#">Element preview</a>
salario	div.col-6 li:nth-of-type(1) > span	SelectorText	no	_root	<a href="#">Element preview</a>
experiencia-minima	div.col-6:nth-of-type(2) li:nth-of-type(2) span	SelectorText	no	_root	<a href="#">Element preview</a>
estudios-minimos	span#prefijoEstMin.list-default-text	SelectorText	no	_root	<a href="#">Element preview</a>
conocimiento1	li:nth-of-type(1) a.tag	SelectorText	no	_root	<a href="#">Element preview</a>
conocimiento2	li:nth-of-type(2) a.tag	SelectorText	no	_root	<a href="#">Element preview</a>

Figura 6. Sitemap con el desglose de elementos de Infojobs en la herramienta Web Scraper

Los mapas de sitio personalizados para cada página se exportan a un archivo en formato JSON

que permite editar el texto e incluir en ellos todas las URL de las páginas de las ofertas. Este paso de edición manual del archivo JSON se puede automatizar si desde las opciones del *sitemap*, al inicio, se especifica el rango de páginas numerado. Para exportar, se accede a esta opción desde el menú de la herramienta *Export Sitemap*, y se copia y pega el código en un editor de textos. El empleado en esta investigación fue el programa *Brackets*, de código abierto, disponible para los sistemas operativos de *Windows* y *MacOs*. En el texto de este archivo, a continuación en `"startUrl":`, se incluyen todas las URL de las ofertas (entrecomilladas y separadas por comas). El listado de URL se puede generar en una hoja de cálculo. Al haber renombrado el archivo html, el proceso se simplifica, ya que todas las direcciones son iguales a excepción del número.

El resultado del archivo JSON es un texto similar al que se muestra a continuación, correspondiente a LinkedIn:

```
{ "_id": "nombredelsitemap", "startUrl": ["http://nombrededominio.com/0001linkedin.html", "http://nombrededominio.com/0002linkedin.html", "http://nombrededominio.com/0003linkedin.html", "http://nombrededominio.com/0004linkedin.html", "http://nombrededominio.com/0005linkedin.html"], "selectors": [{"delay": 0, "id": "Titulo", "multiple": false, "parentSelectors": ["_root"], "regex": "", "selector": "h1", "type": "SelectorText"}] }
```

Una vez modificado el archivo con todas las direcciones incluidas, se importa desde la opción del programa *Import Sitemap* dentro del menú *Create new sitemap*, copiando y pegando el texto del archivo JSON y añadiendo un nombre al mapa del sitio. Con este archivo, el programa ya dispone tanto del listado de páginas de las cuales se quiere extraer la información como de los diferentes campos del contenido del que debe extraer.

La última etapa es lanzar la acción de *scraping* desde el menú del sitemap: *Scrape*. Esta opción, abre el panel desde donde personalizar el intervalo de solicitud, *Request interval*, y el tiempo de carga de la página, *Page load delay*, con el que poder modificar lo que se tarda en realizar la petición a cada página y lo que tarda la página en aparecer. A continuación, al pinchar en el botón *Start Scraping* empieza el proceso automático mediante el cual las páginas se van abriendo de una a una y el sistema se encarga de ir extrayendo los datos de las mismas. Una vez finalizado, aparece una notificación y ya se pueden exportar los datos desde el menú: *Export Data*, con las opciones de descargar el archivo en formato .xlsx o .csv, ambos editables desde un programa de hojas de cálculo. Cabe recordar que hay que realizar este paso para cada *sitemap*



elaborado: uno para LinkedIn y otro para Infojobs, al tratarse de fuentes de datos distintas.

El resultado de cada proceso de *scraping* es una hoja de cálculo donde cada fila contiene todos los datos de una página (oferta) y cada columna contiene cada uno de los selectores configurados (campos de contenido). De esta manera se dispone de todos los datos ordenados y listos para su codificación, filtrado y análisis posterior.

Parte de los hallazgos obtenidos gracias a esta metodología se muestran en el estudio de Escandell-Poveda et al. (2022) en el que se presentan las competencias de un/a especialista SEO. En él, las autoras exponen los requerimientos necesarios para esta figura desde la perspectiva del mundo laboral, concluyendo tanto una serie de saberes y habilidades especializadas como un conjunto de destrezas transversales, entre otras cuestiones. De igual manera, los resultados generados a partir de la aplicación de esta metodología, sirven como punto de partida de la investigación sobre competencias profesionales del perfil SEO (Escandell-Poveda et al., 2023), que combina un enfoque cualitativo y cuantitativo, empleando un diseño exploratorio secuencial (Dexplos). Estos estudios demuestran la eficacia del método y su aplicación empírica para obtener datos provenientes del mercado de trabajo.

#### 4.3. Relevancia y precisión de la búsqueda

En este proceso, se han tenido en cuenta los dos criterios básicos de una búsqueda: la relevancia y la precisión.

La relevancia de la búsqueda está estrechamente relacionada con la decisión sobre las palabras clave utilizadas. Como término de búsqueda, en un primer momento se planteó incluir: “posicionamiento web” o “posicionamiento en buscadores”, por ser expresiones equivalentes al SEO en idioma español. Sin embargo, el pretest aplicado a las primeras ofertas y tras realizar varias búsquedas, se descartó debido a que, en todas las ofertas encontradas a partir de estos términos, aparecía la palabra “SEO”, no ocurriendo lo mismo al contrario.

Se asumió mayor ruido con el propósito de garantizar la recuperación de todos los documentos relevantes, por lo que se debe suponer que se han recuperado todas las ofertas pretendidas bajo la única limitación del término SEO en su contenido. Al prevalecer este criterio, se puede observar la respuesta de la fuente ante búsquedas con procesos automatizados a través del dato de precisión. Así, para conocer la precisión, se revisaron manualmente todas las ofertas recuperadas válidas y se contabilizaron, finalmente,

aquellas que versaban únicamente sobre el perfil del SEO. Este índice no supera el 10% (tabla III), dato que refleja principalmente el efecto que producen las posibilidades de introducción y recuperación de la información de los portales.

	LinkedIn	Infojobs
N (válida): 6.044	4.667	1.377
Ítems sobre el perfil o la competencia SEO: 506	367	139
Relevancia (según criterio): 100 %	100 %	100 %
Precisión: 8,4%	7,9 %	10,1 %

Tabla III. Principales datos para la evaluación de la búsqueda realizada

### 3. Discusión y conclusiones

En este artículo se ha expuesto el diseño del marco metodológico orientado a investigaciones que se basen en el contenido de las ofertas de empleo publicadas en el medio digital. Gracias a su desarrollo empírico, que combina fases manuales y automatizadas, se ha conseguido generar un procedimiento completo aplicable al estudio de características y requerimientos de la demanda laboral extrapolable al estudio de las necesidades sobre un perfil o competencia profesional.

De los resultados de esta investigación se concluye que las técnicas digitales utilizadas son útiles para la extracción de datos, pues las herramientas aplicadas en concreto a este caso, *Link Klipper* y *Web Scraper* han facilitado la recolección, extracción y volcado del contenido de las ofertas de trabajo, tal como se había planificado, facilitando el posterior análisis de los casos de la muestra extraída.

Durante la primera fase del proceso descrito, los recursos aplicados permiten la automatización de las descargas y presentan opciones que facilitan el tratamiento posterior de la información. A este respecto, se destaca el filtro de eliminación de los resultados duplicados que ayuda a la depuración de la información y, por lo tanto, a la conformación de la muestra de estudio. La herramienta que aplica la técnica del *scraping*, en la segunda fase, también presenta ventajas para la extracción de los contenidos y su posterior codificación, pues permiten agilizar la recopilación de todo el contenido recuperado en una hoja de cálculo, con el consiguiente ahorro de tiempo, reducción de esfuerzo por parte de los/as investigadores/as y posibles errores que supondría el hecho de hacerlo manualmente (Niu et al., 2021).

Sin embargo, en la línea de lo escrito por otros autores, se confirma que tales recursos y herramientas aún no pueden suplir la toma de decisiones estratégicas sobre el método, incluso en cuestiones potencialmente automatizables, siempre que puedan llegar a afectar a las evidencias de la investigación. Este es el caso de la selección de las palabras clave a utilizar en la búsqueda. Se trata de uno de los momentos críticos del procedimiento descrito, por lo que es conveniente aplicar alguna prueba inicial para aproximarse a conocer la cadena óptima. En este caso, se optó por utilizar el término "SEO", dado que otras fórmulas más ajustadas al perfil profesional no proporcionaban resultados tan precisos. Esta circunstancia indica, por un lado, que la demanda de trabajo está menos familiarizada con otros términos que se ajustarían mejor al perfil solicitado; y, por otro, que las bases de datos utilizadas no solo son oportunas para el tipo de información que se pretende extraer, además la presentan lo suficientemente estructurada para que un término tan general como es "SEO" sea el adecuado.

Asimismo, una vez seleccionada la muestra, con respecto a la técnica del *scraping* o "raspado", es importante preparar la extracción del contenido proporcionando todos los condicionantes técnicos para ello. En este caso, se ha utilizado un alojamiento web propio y se han guardado y renombrado las páginas para proceder a su rastreo a través de una herramienta específica. Todo ello implica tener competencias especializadas en el entorno digital, que complementan los conocimientos de las técnicas de investigación social.

Además, se han encontrado límites en el empeño por aplicar el proceso automatizado de recuperación de la información, que se pueden atribuir tanto a la fuente como a las diferentes herramientas utilizadas. En concreto, con respecto a la fuente, se aprecia el inconveniente de la visualización de las páginas con usuario registrado en LinkedIn. La estrategia propuesta para este caso supera esta limitación, pero se debe asumir una cadencia temporal entre descargas múltiples de 30 minutos. De igual forma, si la búsqueda empleada arroja un gran volumen de casos, los filtros deben aplicarse utilizando otros complementarios.

En consecuencia, el éxito del proceso automatizado de recuperación de la información depende, en gran medida, de los portales o plataformas de las que se extraen los datos, es decir, de la fuente; aunque las limitaciones detectadas provenientes de las mismas también deberían ser consideradas por las herramientas empleadas, en tanto propuestas de mejora para aumentar sus funcionalidades y eficiencia. El proceso presentado en este artículo es un ejemplo que

puede ser replicado por otros estudios, en cambio también muestra la urgencia de tener instrumentos más potentes o con funciones concretas para la investigación social.

La gran cantidad de información que ofrece internet no tiene parangón. Presenta una oportunidad única para la investigación social, pero cumplir los condicionantes para que pueda ser explotada por los científicos sociales también es una necesidad. Centrándonos en el caso tratado, el estudio de las ofertas de trabajo a través de las plataformas digitales puede proporcionar claves para detectar nichos y tendencias del mercado laboral y, como consecuencia, también contribuye a identificar áreas de mejora en la formación de las diferentes generaciones. Sin embargo, su naturaleza efímera obliga a disponer de sistemas rápidos y eficientes de captura de información, que puedan competir con el volumen y la velocidad que caracteriza la producción de datos de internet, por un lado, y que puedan ser revisados y analizados aplicando las premisas que sostiene el método científico social, por las personas que tienen tales competencias, por otro. El desarrollo de las técnicas digitales en el ámbito de la investigación social y la preparación de sus investigadores es, como consecuencia, un tema urgente que precisa ser debatido por la comunidad científica.

## Agradecimientos

Este trabajo forma parte del proyecto "Parámetros y estrategias para incrementar la relevancia de los medios y la comunicación digital en la sociedad: curación, visualización y visibilidad (CUVICOM)". PID2021-123579OB-I00 (MICINN), Ministerio de Ciencia e Innovación (España).

## Referencias

- Álvarez-Flores, Erika-Patricia; Núñez-Gómez, Patricia; Olivares-Santamarina, Jose-P. (2018). Perfiles profesionales y salidas laborales para graduados en Publicidad y Relaciones públicas: de la especialización a la hibridación. // *Profesional de la Información*. 27:1, 136-147. <https://doi.org/10.3145/epi.2018.ene.13> (2018-12-12).
- Arcila-Calderón, Carlos; Van Atteveldt, Wouter; Trilling, Damian (2021). Special Issue Computational Methods and Big Data in Communication Research. // *Cuadernos. info*. 49, I-IV. <https://doi.org/10.7764/cdi.49.35333> (2021-12-18).
- Baeza-Servin, Mauricio (2015). Documentación de EasyDev. <https://easydev.readthedocs.io/es/latest/#> (2019-01-15).
- Baeza-Servin, Mauricio (2015, diciembre 3). Re: Unir varios archivos calc en uno solo. // Foro Oficial de Apache OpenOffice. <https://forum.openoffice.org/es/forum/viewtopic.php?f=21&t=11598> (2019-01-15).
- Blog Open-Office (2015, diciembre 5). EasyDev o la extensión en Python que te facilitará la vida en Basic. <https://blog.open-office.es/extensiones/easydev-extension-python-que-te-facilitara-la-vida-en-basic> (2018-11-20).

- Boettger, Ryan-K.; Lam, Chris; Palmer, Laura (2017). Improving the data information literacies of technical communication undergraduates. // 2017 IEEE International Professional Communication Conference (ProComm). IEEE. 1-5. <https://doi.org/10.1109/IPCC.2017.8013934> (2021-12-18).
- Clemente-Mediavilla, Jorge; Antolín-Prieto, Rebeca (2019). LinkedIn job offers aimed at Advertising graduates in Spain. // Profesional de la Información. 28:4. <https://doi.org/10.3145/epi.2019.nov.13> (2019-12-20).
- Dewi, Lusiana Citra; Chandra, Alvin-Chandra (2019). Social media web scraping using social media developers API and regex. // Procedia Computer Science. 157, 444-449. <https://doi.org/10.1016/j.procs.2019.08.237> (2021-12-18).
- Escandell-Poveda, Raquel; Papi-Gálvez, Natalia; Iglesias-García, Mar (2022). Competences of SEO specialist: a perspective from the labor market. Technical note. // Profesional de la Información. 31:5, e310513 <https://doi.org/10.3145/epi.2022.sep.13>
- Escandell-Poveda, Raquel; Papi-Gálvez, Natalia; Iglesias-García, Mar (2023). Competencias profesionales en perfiles digitales: especialistas en posicionamiento web. // Revista de Comunicación. 22:1, 109-125. <https://doi.org/10.26441/RC22.1-2023-3034>
- Federico. (2017, enero 27). Re: macro para juntar varias hojas en una. // Foro oficial de Apache OpenOffice. <https://forum.openoffice.org/es/forum/viewtopic.php?f=21&t=6232#p23134> (2019-01-15).
- Flores-Márquez, Dorismilda; González-Reyes, Rodrigo (2021). En busca de coordenadas metodológicas para estudiar la cultura digital. // Flores-Márquez, Dorismilda; González Reyes, Rodrigo (eds.). La imaginación metodológica: Coordenadas, rutas y apuestas para el estudio de la cultura digital. Tintable. 15-24.
- Gandomi, Amir; Haider, Murtaza (2015). Beyond the hype: Big data concepts, methods, and analytics. // International journal of information management. 35:2, 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007> (2019-12-20).
- Giomelakis, Dimitrios; Karypidou, Christina; Veglis, Andreas (2019). SEO inside Newsrooms: Reports from the Field. // Future internet. 11:12, 261. <https://doi.org/10.3390/fi11120261> (2020-11-01).
- Hillen, Judith (2019). Web scraping for food price research. // British Food Journal. 121:12, 3350-3361. <https://doi.org/10.1108/BFJ-02-2019-0081> (2020-11-01).
- Infoempleo; EY. (2018). Informe 2018. talento conectado: Nuevas realidades del mercado de trabajo en España. [https://cdn.infoempleo.com/infoempleo/documentacion/publicaciones/Informe\\_talento\\_conectado.pdf](https://cdn.infoempleo.com/infoempleo/documentacion/publicaciones/Informe_talento_conectado.pdf) (2019-01-30).
- Infojobs (2022). Casi 6 de cada 10 empresas consultan las redes sociales de una persona antes de contratarla, 10 puntos porcentuales más que en 2021. // Informe Redes Sociales y Marca Personal Infojobs. <https://acortar.link/B3WGaN> (2022-07-01).
- Jiménez-Gómez, Isidro; Mañas-Viniestra, Luis (2018). Ofertas de empleo mejor remuneradas en comunicación y diseño: nuevos perfiles y efecto full-stack. // Doxa Comunicación. 27, 239-251. <https://doi.org/10.31921/doxacom.n27a12> (2019-12-20).
- Lafrate, Fernando (2014). A Journey from Big Data to Smart Data. // Benghozi, Pierre-Jean; Krob, Daniel; Lonjon, Antoine; Panetto, Hervé (eds.). Digital Enterprise Design & Management. Advances in Intelligent Systems and Computing. 261, Springer, Cham. [https://doi.org/10.1007/978-3-319-04313-5\\_3](https://doi.org/10.1007/978-3-319-04313-5_3) (2021-06-05)
- Laney, Doug (2001, febrero 6). 3D Data Management: Controlling Data Volume, Velocity, and Variety. // Application Delivery Strategies. Meta Group. <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an...> (2022-06-01).
- Lasén, Amparo; Puente, Héctor (2016). La cultura digital. Universitat Oberta de Catalunya. <https://acortar.link/bHXHcz> (2021-06-05).
- Leonelli, Sabina (2020). Scientific Research and Big Data. // Zalta, Edward N. (ed.), The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/sum2020/entries/science-big-data> (2022-06-01).
- Lopezosa, Carlos; Codina, Lluís; Díaz-Noci, Javier; Ontalba-Ruipérez, Jose A. (2020). SEO and the digital news media: From the workplace to the classroom. // Comunicar. 28:63, 63-72. <https://doi.org/10.3916/C63-2020-06> (2021-06-05).
- Marres, Noortje; Weltevrede, Esther (2013). Scraping the social? Issues in live social research. // Journal of Cultural Economy. 6:3, 313-335. <https://doi.org/10.1080/17530350.2013.772070>
- Marta-Lazo, Carmen; González Aldea, Patricia; Herrero Curiel, Eva (2018). Professional Skills and Profiles in Journalism Demanded by Companies: Analysis of Offers at LinkedIn and Infojobs. // Communication & Society. 31:4, 211-228. <https://doi.org/10.15581/003.31.4.211-227> (2021-06-05).
- Massimino, Brett (2016). Accessing online data: Web-crawling and information-scraping techniques to automate the assembly of research data. // Journal of Business Logistics. 37:1, 34-42. <https://doi.org/10.1111/JBL.12120> (2022-06-05)
- Miguel-San-Emeterio, Begoña (2018). Las competencias digitales en los grados de periodismo, publicidad y relaciones públicas y comunicación audiovisual en la universidad: el caso de la Comunidad de Madrid. (Tesis doctoral). Universidad Complutense de Madrid, E-prints Complutense. <https://eprints.ucm.es/id/eprint/50690/>
- Milovanović, Stevan; Bogdanović, Zorica; Labus, Aleksandra; Despotović-Zrakić, Marijana; Mitrović, Svetlana (2022). Social recruiting: An application of social network analysis for preselection of candidates. // Data Technologies and Applications. 56:4, 536-557. <https://doi.org/10.1108/DTA-01-2021-0021>
- Ministerio de Asuntos Económicos y Transformación Digital (2021). España digital 2025. [https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/230720-Espa%C3%B1aDigital\\_2025.pdf](https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/230720-Espa%C3%B1aDigital_2025.pdf) (2022-06-01).
- Moo-Herrera, Armando; Aranda, Ramón; Hernández-López, Francisco-Javier (2020). Herramientas Web Scraping para la Adquisición de Datos Turísticos. // KIKAME. 9:9, 100-111.
- Morales-i-Gras, Jordi; Sánchez-i-Vallès, Oriol (2022). El 14F a Instagram. Una proposta d'articulació de tècniques de raspateo i anàlisi de xarxes. // Papers. Revista de Sociologia. 107:1, 147-174. <https://doi.org/10.5565/rev/papers.2967> (2022-08-09).
- Morato, Jorge; Sánchez-Cuadrado, Sonia; Fernández-Bajón, María-Teresa (2016). Tendencias en el perfil tecnológico del profesional de la información. // Profesional de la Información. 25:2, 169-178. <http://dx.doi.org/10.3145/epi.2016.mar.03> (2019-12-20).
- Niu, Qingli; Kandhro, Irfan-Ali; Kumar, Anil; Shah, Shahna-waz; Hasan, Muhammad; Ahmed, Hifza-Mehfooz; Liang, Fei (2022). Web Scraping Tool For Newspapers And Images Data Using Jsonify. // Journal of Applied Science and Engineering. 26:4, 465-474. [https://doi.org/10.6180/jase.202304\\_26\(4\).0002](https://doi.org/10.6180/jase.202304_26(4).0002)
- Pereira, Renita-Crystal; Vanitha, T. (2015). Web scraping of social networks. // International Journal Innovation Research Computer Communication Engineering. 3:7, 237-240. [https://www.ijrcce.com/special-issues/pdf/2015/october/43\\_710.pdf](https://www.ijrcce.com/special-issues/pdf/2015/october/43_710.pdf) (2021-05-20).

- Popescu, Cristian-Aurelian; Simion, Cristina-Petronela; Popescu, Ana-María (2019). Digital marketing in Romania: An overview. // Proceedings of the 33rd International Business Information Management Association Conference, IBIMA 2019: Education Excellence and Innovation Management through Vision 2020. 5377-5381.
- Rodríguez Cano, César-Augusto (2022). Hipermétodos. Repertorios de la investigación social en entornos digitales. // UAM, Unidad Cuajimalpa, Ciencias de la Comunicación y Diseño. <https://doi.org/10.24275/9786072824812> (2022-09-06).
- Rogers, Richard (2013). Digital methods. // The MIT Press. <https://doi.org/10.7551/mitpress/8718.001.0001> (2021-05-20).
- Salcedo Maldonado, Jose-Luís; Fuster Morell, Mayo (2014). Métodos de investigación en la red. // Equipo IGOPnet (ed.), Jóvenes, Internet y Política. Centro Reina Sofía sobre Adolescencia y Juventud Fundación de Ayuda contra la Drogadicción (FAD). 135-172. [https://ddd.uab.cat/pub/caplli/2014/190272/jovintpol\\_a2014p135iSPA.pdf](https://ddd.uab.cat/pub/caplli/2014/190272/jovintpol_a2014p135iSPA.pdf) (2022-06-05).
- Saura, José-Ramón; Palos-Sánchez, Pedro R.; López, Juan-José (2017). The digital marketing environment through search engine optimization and web metrics. // Proceedings of the International Conference on WWW/Internet 2017 and Applied Computing 2017. 307-309.
- Sheffield, Jenna Pack (2020). Search engine optimization and business communication instruction: interviews with experts. // Business and professional communication quarterly. 83:2, 153-183. <https://doi.org/10.1177/2329490619890335> (2022-06-07).
- Sued, Gabriela-Elisa (2020). Repertorio de técnicas digitales para la investigación con contenidos generados en redes sociodigitales. // PAAKAT: revista de tecnología y sociedad. 10:19. <https://doi.org/10.32870/pk.a10n19.498> (2022-10-09).
- The Social Media Family (2017). Primer monográfico sobre los procesos de selección en competencias digitales. <https://thesocialmediafamily.com/news-posts/monografico-competencias-digitales/> (2021-05-20).
- TMP Worldwide Spain (2017). Empleómetro 2017. Estudio de inversión y publicación de ofertas en jobsites. // WTC Spain. (2022-06-05).
- Valle, Ángel; De-la-Peña, Carlos (2014). Del dato a la toma de decisiones, pasando por el conocimiento. // Harvard Deusto Márketing y Ventas, 127, 34-41.
- Van Rijsbergen, Cornelis-Joost (1979). Chapter Seven: Evaluation. // Van Rijsbergen, Cornelis-Joost (Ed.), Information Retrieval. Butterworths. 112-140 <https://www.dcs.gla.ac.uk/Keith/pdf/Chapter7.pdf> (2021-06-05)
- Wang, Hui-Ju (2022). Market segmentation of online reviews: A network analysis approach. // International Journal of Market Research. 64:5, 652-671. <https://doi.org/10.1177/14707853211059076> (2022-10-09).
- WiecheteK, ŁuKasz; Pastuszek, Zbigniew (2022). Academic social networks metrics: an effective indicator for university performance? // Scientometrics. 127:3, 1381-1401. <https://doi.org/10.1007/s11192-021-04258-6> (2022-10-09).
- Zarrabeitia-Bilbao, Enara; Morales-i-Gras, Jordi; Rio-Belver, Rosa María; Garechana-Anacabe, Gaizka (2022). Green energy: identifying development trends in society using Twitter data mining to make strategic decisions. // Profesional de la Información. 31:1. <https://doi.org/10.3145/epi.2022.ene.14> (2022-10-09).

---

Enviado: 2022-12-03. Segunda versión: 2023-06-02.  
Aceptado: 2023-06-08.

---