

SCIRE

Representación y organización del conocimiento

SCIRE

Representación y organización del conocimiento

Vol. 29, n.º 1, enero-junio 2023

ISSN 1135-3716

ISSN (e) 2340-7042

Scire:

knowledge representation and organization
Vol. 29, n. 1, January-June 2023

Ibersid:

Red de Investigación
en Sistemas de Información
y Documentación

© 2023 Los autores y autoras conservan sus derechos de autor, aunque ceden a la revista de forma no exclusiva los derechos de explotación (reproducción, distribución, comunicación pública y transformación) y garantizan a esta el derecho de primera publicación de su trabajo, el cual estará simultáneamente sujeto a la licencia CC BY-NC-ND. Los autores aceptan la responsabilidad legal de cumplir plenamente con los códigos éticos y leyes apropiadas, y de obtener todos los permisos de derecho de autor debidos. Se permite y se anima a los autores y autoras a difundir electrónicamente la versión editorial (versión publicada por la editorial) en la página web personal del autor y en el repositorio de la institución a la que pertenece.

ISSN: 1135-3716 = Scire (Zaragoza)

ISSN (e): 2340-7042

Depósito legal: Z. 1.790 — 1995

Edita: Ibersid® con la colaboración de Prensas de la Universidad de Zaragoza

Imprime:

Servicio de Publicaciones. Universidad de Zaragoza.

Edificio de Ciencias Geológicas, C/ Pedro Cerbuna, 12.

50009 Zaragoza, España. Tel.: 976 761 330. Fax: 976 761 063.

Scire

representación y organización
del conocimiento

Alcance y objetivos

Scire: representación y Organización del Conocimiento es una publicación semestral de carácter interdisciplinar sobre la representación, normalización, tratamiento, recuperación y comunicación de la información y el conocimiento.

Difusión

Scire tiene difusión internacional. Agradecemos la inclusión en los siguientes servicios de referencia: Scopus, ESCI, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC y Catálogo Latindex.

Instrucciones para los autores y procedimiento de evaluación

La última versión de las instrucciones para presentación de trabajos y del procedimiento de evaluación editorial están disponibles en: <https://www.iberlid.eu/ojs/index.php/scire/about/submissions>

Agradecimientos

Agradecemos el apoyo del Departamento de Ciencia, Universidad y Sociedad del Conocimiento del Gobierno de Aragón con su subvención a grupos de investigación S6520D, al Vicerrectorado de Investigación y a la Facultad de Filosofía y Letras de la Universidad de Zaragoza.

Redacción, distribución y canje

Revista Scire
Departamento de Ciencias de la Documentación e Historia de la Ciencia
Facultad de Filosofía y Letras
Universidad de Zaragoza
C/ Pedro Cerbuna 12,
E-50.009 Zaragoza (Spain)

Tfno: int + 34 976 762239. Fax: 34 976761506.
E-mail: mailto:scire@unizar.es

Suscripciones y números sueltos

Suscripción anual: 30 €. Número suelto: 20 €. (IVA inc.)

Scire

knowledge organization
and representation

Scope and aims

Scire: Representación y Organización del Conocimiento is an interdisciplinary journal published twice a year that is devoted to the representation, standardization, treatment, retrieval and communication of information and knowledge.

Dissemination

Scire has international distribution. We acknowledge its inclusion in the following reference services: Scopus, ESCI, Information Science Abstracts, Information Services in Physics, Electronics and Computing, Library and Information Science Abstracts, Sociological Abstracts, ERIH Plus, Knowledge Organization Literature, Base de Datos ISOC and Catálogo Latindex.

Instructions for authors and evaluation process

The last version of the instructions for authors and assessment process is available at: <https://www.iberlid.eu/ojs/index.php/scire/about/submissions>

Acknowledgments

We acknowledge the help of the Department of Science, University and Knowledge Society of the Government of Aragón (grant S6520D to research groups), and of the Research Vice Rectorate and the Faculty of Philosophy and Arts of the University of Zaragoza.

Contact address

Revista Scire
Departamento de Ciencias de la Documentación e Historia de la Ciencia
Facultad de Filosofía y Letras
Universidad de Zaragoza
C/ Pedro Cerbuna 12,
E-50.009 Zaragoza (Spain)

Tel.: int + 34 976 762239. Fax: 34 976761506.
E-mail: scire@unizar.es

Subscriptions

Annual subscription: 30 €. Issue: 20 €. (VAT included)

Editor

Francisco Javier García Marco, Univ. de Zaragoza. E-mail: jgarcia@unizar.es

Consejo de redacción / Editorial council

Mario Guido Barité Roqueta,
Universidad de La República, Uruguay

José Augusto Chaves Guimarães,
Universidade Estadual Paulista, Brasil

João Batista Ernesto Moraes,
Universidade Estadual Paulista, Brasil

Francisco Javier García Marco,
Universidad de Zaragoza, España

Daniel Martínez Ávila,
Universidad de León, España

Francisco Javier Martínez Mendez,
Universidad de Murcia, España

Álvaro Quijano Solís,
Colegio de México, México

Consejo científico / Scientific council

Isidro Aguillo Caño, IPP-CSIC, España

Tomás Baiget, EPI S. A., España

José Luis Bonal Zazo, Univ. de
Extremadura, España

Mercedes Caridad Sebastián,
Universidad Carlos III de Madrid, España

Alberto Carreras Gargallo,
Universidad de Zaragoza, España

Constança Espelt Busquets,
Universidad de Barcelona, España

Juan Carlos Fernández Molina,
Univ. de Granada, España

María Eulalia Fuentes Pujol, Universidad
Autónoma de Barcelona, España

Fernando Galindo Ayuda,
Universidad de Zaragoza, España

Blanca Gil Urdiciáin, Universidad
Complutense de Madrid, España

Isidoro Gil Leiva,
Universidad de Murcia, España

Alan Gilchrist, Cura Consortium,
Reino Unido

Vicente Pablo Guerrero Bote, Universidad
de Extremadura, España

Víctor Herrero Solana,
Univ. de Granada, España

José María Izquierdo Arroyo,
Universidad de Murcia, España

María Pilar Lasala Calleja,
Universidad de Zaragoza, España

Alfonso López Yepes, Universidad
Complutense de Madrid, España

José López Yepes, Universidad
Complutense de Madrid, España

Pedro Marijuán Fernández,
Universidad de Zaragoza, España

Bonifacio Martín del Brío,
Universidad de Zaragoza, España

José Antonio Moreiro González,
Universidad Carlos III de Madrid, España

Purificación Moscoso Castro,
Universidad de Alcalá, España

Félix Moya Anegón,
Universidad de Granada, España

María del Carmen Negrete Gutiérrez,
Universidad Autónoma de México

Catalina Naumis Peña,
Universidad Autónoma de México

José Luis Otal, Universidad Jaume I de
Castellón, España

Manuel José Pedraza Gracia,
Universidad de Zaragoza, España

María Pinto Molina,
Universidad de Granada, España

Gloria Ponjuán Dante,
Universidad de La Habana, Cuba

Blanca Rodríguez Bravo,
Universidad de León, España

José Vicente Rodríguez Muñoz,
Universidad de Murcia, España

Adelaida Román Román,
CINDOC (Madrid), España

Juan Ros García,
Universidad de Murcia, España

Francisco José Ruiz de Mendoza Ibáñez,
Universidad de La Rioja, España

Félix Sagredo Fernández,
Universidad Complutense de Madrid, España

Elías Sanz Casado,
Universidad Carlos III de Madrid, España

Carlos Serrano Cinca,
Universidad de Zaragoza, España

Revisores externos del número / External reviewers in this issue

Agradecemos enormemente la colaboración altruista y desinteresada de Carlos Díaz Redondo, Cristina Faba-Pérez, Mariàngela Spotti Lopes Fujita, Javier Lacasta Miguel, Gonzalo Marco Cuenca, Celso Martínez Musiño, Juan Antonio Pastor Sánchez, Rodrigo Rabello, Merizanda María del Carmen Ramírez Aceves, Ana Belén Ríos Hilario, Jesús Robledano Arillo, José Antonio Salvador Oliván, Juan Miguel Sánchez Vigil, Carlos Tejada Artigas

Candidaturas al consejo científico

Se aceptan candidaturas al consejo científico de especialistas del área de Biblioteconomía y Documentación y de otras disciplinas relacionadas (Informática, Ciencias Sociales, Lingüística, Filosofía, Psicología, etc.) con experiencia profesional e investigadora demostrada. En el sistema público de investigación español, suele ser equivalente al doctorado y dos sexenios de investigación o méritos equivalentes.

Scientific council membership policy

Candidatures of researchers from LIS and other related disciplines (Computer Science, Social Sciences, Linguistics, Philosophy, Psychology, etc.) with demonstrated professional and research experience are welcomed. In the Spanish public research system, for example, this usually means having a doctorate and two scientific productivity sexennia or equivalent outputs.

Tabla de contenidos en español

Table of contents in Spanish

<i>Tabla de contenidos en español</i>	9
<i>Tabla de contenidos en inglés</i>	11

Artículos

Análisis de respuestas enriquecidas en Google

Sonia SANCHEZ-CUADRADO
Jorge MORATO.....13

*Mención a la bibliografía en los metadatos
de representación de objetos museables*

Beatriz TARRÉ ALONSO
Camila MONTEIRO DE BARROS25

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

Raquel ESCANDELL-POVEDA
Natalia PAPI-GÁLVEZ
Mar IGLESIAS-GARCÍA31

*Crítica diplomática: un estudio panorámico
de la producción científica*

Sonia Oliveira Matos MOUTINHO
Sonia Maria TROITIÑO RODRIGUEZ43

*El artefacto como sistema generador
de comunidades epistémicas en ambientes
científicos de la nanotecnología*

Heriberto RAMÍREZ
Javier TARANGO
Juan D. MACHIN-MASTROMATTEO51

Índices

<i>Índice de autores</i>	63
<i>Índice de materias en español</i>	63
<i>Índice de materias en inglés</i>	63

Tabla de contenidos en inglés

Table of contents in English

Table of contents in Spanish9
Table of contents in English11

Articles

Analysis of rich answers in Google.....13
Sonia SANCHEZ-CUADRADO
Jorge MORATO.....13

*Bibliographic notes in the representation
metadata of museum objects*25
Beatriz TARRÉ ALONSO
Camila MONTEIRO DE BARROS25

*Digital techniques for the study of professional
skills and profiles: the case of SEO job offers*31
Raquel ESCANDELL-POVEDA
Natalia PAPI-GÁLVEZ
Mar IGLESIAS-GARCÍA31

*Diplomatic criticism: a panoramic study
of scientific production*43
Sonia Oliveira Matos MOUTINHO
Sonia María TROITIÑO RODRIGUEZ43

*The artefact as a generating system
of epistemic communities in scientific
environments of nano-technology*51
Heriberto RAMÍREZ
Javier TARANGO
Juan D. MACHIN-MASTROMATTEO51

Indexes

Author index63
Subject index in Spanish63
Subject index in English63

Análisis de respuestas enriquecidas en Google

Analysis of rich answers in Google

Sonia SANCHEZ-CUADRADO (1), Jorge MORATO (2)

(1) Universidad Complutense, C/Santísima Trinidad, 37 28010 Madrid, España, sscuadrado@ucm.es

(2) Universidad Carlos III, Avda. Universidad, 30, 28912 Leganés, Madrid, jmorato@inf.ucm.es

Resumen

En recuperación de información web, los motores de búsqueda como Google incluyen funcionalidades que devuelven respuestas directas a las consultas de los usuarios. Estas respuestas tratan de resolver una necesidad informativa y se conocen como *rich answers*. Para determinar cómo se presentan estos resultados y cómo afecta la optimización de los motores de búsqueda, se ha realizado un análisis de las respuestas directas destacadas que presenta Google. En este trabajo se han examinado preguntas de tipo informacional expresadas en lenguaje natural con los términos "what is". Se ha analizado el listado de los resultados para identificar las características de las respuestas directas. Además, se han explorado las estrategias SEO que puedan determinar la relevancia del fragmento respecto de la consulta. Con este trabajo se constata que la respuesta no se extrae necesariamente de forma literal de una página web y se comprueba que la solución a las preguntas puede proceder de varios recursos. Los fragmentos de respuesta directa y otros *rich answers* pueden llegar a ocupar cerca de la mitad de la página principal de resultados, cobrando un mayor protagonismo y desplazando al resto de los resultados orgánicos. Las respuestas directas proporcionan un cambio en los hábitos de búsqueda y un nuevo modo de navegar en la red basado en un sistema hiperenlazado de pregunta respuesta.

Palabras clave: Respuestas enriquecidas. Fragmentos enriquecidos. GAB. SEO. Motores de búsqueda. Google. Optimización de respuestas.

1. Introducción

Los motores de búsqueda son el principal punto de acceso para encontrar y consultar información disponible en Internet. En la actualidad, los buscadores web representan una herramienta imprescindible para los internautas y la recuperación de información. Aunque existe un amplio número de buscadores, Google figura como el buscador con mayor cuota de mercado tanto a nivel mundial (StatCounter, 2023) como en España (StatCounter, 2022). Esto supone un ingente volumen de páginas web y una elevada competencia para disputarse el tráfico. Por eso, los sitios web que desean tener visibilidad aspiran no solo a estar incluidos en el índice del buscador, sino a posicionarse entre los primeros resultados (Killo-ran, 2013). Los buscadores ofrecen el listado de

Abstract

In Web Information Retrieval, search engines, such as Google, include features that return direct responses to user queries. These answers try to solve an information need and are known as rich answers. To find out how these results are presented and how it affects search engine optimization, an analysis of the featured direct responses presented by Google has been carried out. In this work, a collection of informational questions expressed in natural language with the terms "what is" has been examined. The ranking of results has been analysed, identified the characteristics of the direct responses. Also, SEO factors that can determine the relevance of the fragment with respect to the query have been explored. This work confirms that the answer is not necessarily taken literally from the web page and confirms that the solution to the questions can come from various resources. The direct response fragments and other rich answers can occupy close to half of the main results page, taking on a greater role and scroll down the rest of the organic results. Direct answers provide a change in search habits and a new way of browsing the web based on a hyperlinked question-answer system.

Keywords: Rich answers. Featured snippets. GAB. SEO. Search engine. Google. AEO.

documentos ordenados por relevancia basada en decenas de criterios. Aunque se desconoce el peso de cada factor, las páginas web tratan de optimizar el contenido para el motor de búsqueda (SEO) de algunos criterios conocidos, con el fin de conseguir valores altos de relevancia y una posición entre los primeros resultados (Morato et al., 2013; Enge et al., 2015; Harto, 2019; Ziakis, 2019).

En la página principal del motor de búsqueda, también aparecen respuestas precisas a las consultas en un lugar destacado (Miklosik, 2019). Para Google, la primera página de resultados (SERP, *search engine results page*) es esencial y se traduce en continuos cambios en su anatomía y distribución del contenido. La disposición de los elementos de la SERP se modifica en función de

las bases de conocimiento del buscador, la consulta que ejecuta el usuario y las variaciones en el algoritmo de recuperación y posicionamiento. Fruto de las estrategias de posicionamiento y de la navegación vertical aparecen en la SERP resultados procedentes de las pestañas de videos, noticias o *shopping*. Visualmente, este tipo de elementos supone un resultado atrayente para el usuario, porque ocupa un espacio considerable además de proporcionar información detallada (por ejemplo, título, enlace, miniatura). Al final de la primera página de resultados también se ubica un bloque para búsquedas relacionadas con consultas de otros usuarios. Se muestran unos ocho términos compuestos de búsqueda con las palabras clave de la consulta ubicados en dos columnas.

Otro de los cambios significativos es la incorporación de respuestas directas (*rich answers*) que proporcionan una contestación breve a una necesidad de información concreta. Estas respuestas adoptan formas distintas en función de la fuente. Sin embargo, se desconocen los criterios por los que estos recursos son seleccionados para proporcionar la respuesta.

El propósito de este estudio es analizar las respuestas directas destacadas que presenta Google en la SERP. Se ejecutan una batería de consultas informacionales y se exploran los resultados en busca de factores que contribuyan a determinar los criterios de relevancia. De acuerdo con este propósito general de la investigación, se formulan los siguientes objetivos: 1) descubrir los tipos de respuestas enriquecidas que aparecen en la SERP; 2) explorar los resultados de las respuestas enriquecidas; y 3) determinar los factores de relevancia para las respuestas enriquecidas.

2. Consultas y respuestas en Google

2.1. Tipo de consultas

Las consultas pueden ser de varios tipos según los datos requeridos (Moldovan et al., 2000; Broder, 2002; Mishra & Jain, 2016) y precisan de un proceso concreto para ser resueltas. No existe un único criterio para categorizar la necesidad del usuario y las respuestas del buscador, pero los primeros trabajos determinaron tres tipos esenciales por la intención de la consulta (navegacionales, transaccionales, informacionales) (Broder, 2002). Otras propuestas se centran en el criterio de clasificación sobre el tipo de información buscada (p.e. objetiva, de opinión, noticias) (Lewandowsky, et al., 2012) o el tipo de respuesta proporcionada, ya sea breve o extensa (Khashabi, et al., 2021).

Las consultas transaccionales (por ejemplo, escuchar música, jugar o descargar un archivo,

comprar un smartphone, o vuelos baratos a...) interesan especialmente por la rentabilidad económica. Disponen incluso de un espacio adicional (*shopping*) dirigido a la actividad comercial. Los resultados de las consultas transaccionales suelen optimizarse con SEO y marketing en motores de búsqueda (SEM).

Las consultas informacionales proporcionan resultados para satisfacer una necesidad de información del usuario, ya sea factual, procedimental o conceptual. El desarrollo tecnológico de los últimos años ha favorecido el desarrollo de preguntas completas en lenguaje natural para traducir la necesidad del usuario. Este proceso de comunicación con la computadora ha originado la ejecución de preguntas completas que incorporan los términos "*what is / qué es*" o "*how to / cómo ...?*" a la frase que define la consulta.

2.2. Tipo de respuestas directas

Google, en su faceta de sistema pregunta-respuesta, trata de devolver resultados en forma de respuesta enriquecidas (*rich answers*) para resolver una necesidad informativa. Entre los tipos de respuestas enriquecidas presentadas en la SERP se encuentran los basados en imágenes, elementos gráficos, texto, clasificación en pestañas, diapositivas, tablas o resultados de formularios. Para una consulta pueden aparecer distintas respuestas enriquecidas combinadas, ya sean visualizadas al mismo tiempo u organizadas en pestañas, diapositivas o carruseles de imágenes (Zhao et al., 2019). También se utiliza un espacio destacado en la SERP denominado *Google Answer Box* (GAB) para respuestas concisas a la consulta de un usuario (Miklosik et al., 2016; Sullivan, 2018; Miklosik, 2019; Khashabi et al., 2021). Este espacio destacado se ubica en la posición cero. Esta posición supone un lugar privilegiado en la parte superior de la página que asegura la visualización e incremento del tráfico para el recurso. Se materializa con los fragmentos destacados (*feature answers*) a modo de respuestas directas enriquecidas (Strzelecki & Rutecka, 2019). Estos resultados directos ofrecen al usuario una solución de consulta sin necesidad de clicar (Ishkin, 2019).

2.2.1. *Featured answer*

Una respuesta destacada o *featured answer* es un fragmento o resumen extraído del texto de una página web que responden de forma directa a una consulta ejecutada en el buscador (Miklosik, 2019; Strzelecki & Rutecka, 2020a). Los fragmentos destacados tratan de proporcionar una respuesta completa ya sea simple o con varios pasos o piezas de información (por ejemplo, respuestas

a *how to?*). Los mecanismos para comprender la consulta del usuario y ofrecer la respuesta precisa presentan cierto grado de sofisticación (Nayak, 2019; Nayak, 2022). Se visualizan como párrafo, listados, tablas, gráficos, imágenes o una combinación de estos elementos. Supone un recurso para acelerar el proceso de resolver la necesidad de información, permitiendo optimizar la experiencia del usuario (Strzelecki & Rutecka, 2020b).

2.2.2. Panel de conocimientos

Otro tipo de respuestas destacadas son las tarjetas de conocimiento que aparecen en el lateral derecho de la SERP enmarcadas en un cuadrado. Muestra información a modo de resumen sobre la consulta. Esta información procede del propio buscador o fuentes de conocimiento como la Wikipedia y páginas de instituciones públicas o privadas. Estos paneles también están relacionados con los gráficos de conocimiento (Singhal, 2012) y con la información local de las ubicaciones de negocios para su representación en el mapa. En ocasiones aparecen *widgets* con alguna funcionalidad como realizar una traducción, buscar un término, o efectuar una conversión de unidades.

2.2.3. Rich snippet

Los fragmentos enriquecidos o *rich snippets* muestran una información destacada del contenido de la página web. Suelen ofrecer una respuesta concisa a la consulta de un usuario en forma de fragmento o párrafo marcado previamente. Se utilizan datos estructurados dentro del código fuente para clasificar la información o expresar una función (Chen et al., 2018). Asimismo, se proporciona metainformación sobre el recurso y el contenido de la página web. Los datos se presentan en forma de resumen, reseñas, valoraciones, eventos, etc. de acuerdo con la clasificación de schema.org (Enge et al., 2015; Zhao et al., 2019).

2.2.4. Otras preguntas de los usuarios

También aparece en la página de resultados un bloque denominado "Otras preguntas de los usuarios" conocidas como PAA del inglés *People Also Ask*. Este bloque recoge sugerencias de consultas en forma de pregunta autogeneradas. Las preguntas se expresan en lenguaje natural y comienzan por palabras interrogativas como qué, quién, cuál, cómo, por qué y dónde, principalmente. Tiene un formato dinámico y recursivo donde al desplegar una pregunta para obtener la respuesta se origina otro bloque con nuevas preguntas de otros usuarios relacionadas. Representan técnicas de navegación y redirección de resultados, pero también iniciativas para afinar la

experiencia del usuario al estilo de las propuestas alternativas a la consulta "Sugerencias de Google".

2.3. Selección de rich answers

En todos estos tipos de respuesta, el algoritmo de Google decide los criterios para mostrar una *featured answer* o una PAA (Sullivan, 2018; Nayak, 2019; Nayak, 2020). Las investigaciones, que tratan de averiguar cómo se seleccionan y extraen las respuestas directas, apuestan por criterios como la velocidad de carga de las páginas, las palabras clave, la autoridad del dominio de la página o la longitud de las consultas (Miklošik et al., 2016).

2.4. Impacto de la rich answer en la SERP

Los resultados enriquecidos, ya sean fragmentos directos o con formato visual, repercuten de forma positiva en la experiencia del usuario (Yu, 2014: 492), mejorando la accesibilidad y usabilidad de la SERP. Respecto al buscador, la percepción del usuario mejora con los resultados directos pues el acceso directo permite descartar cualquier otro resultado orgánico. Desde el punto de vista del recurso, las páginas web experimentan una mayor visibilidad y un aumento del tráfico (Strzelecki & Rutecka, 2020a; Duong, 2019), contribuyendo a mejorar su autoridad del dominio. Desde el punto de vista del usuario, algunos trabajos han estudiado cómo las respuestas enriquecidas condicionan a los internautas quienes tienden a sobrevalorar la confianza en un sitio web (Bernstein et al., 2012; Lurie & Mustafaraj, 2018). Con estos estudios los autores afirman que las respuestas enriquecidas alteran el comportamiento informacional de los usuarios (Bilal & Huang, 2019; Wu, et al., 2020; Bink et al., 2022).

Las respuestas enriquecidas también han contribuido al cambio en los hábitos de la búsqueda de información a través de los asistentes de voz (Trippas, 2020). Esta opción permite formular consultas de forma verbal a través de las funciones del habla en sistemas como Google Search Assistant, Siri o Alexa. La respuesta que proporciona el asistente de voz procede de los resultados directos (*rich answers, google answer box, feature snippet...*). Este único resultado puede ser escuchado en los asistentes de voz (Epstein et al., 2022), pero se omiten los demás. El uso de los asistentes de voz para demandar información se ha extendido con rapidez. Algunos estudios estiman un incremento de más del 20 % en servicios de búsqueda por voz (Mander & Buckle, 2018, eMaketer & Insider Intelligence, 2022) por lo tanto también aumenta la repercusión de las respuestas directas.

Ante tal aumento de oportunidades, cobra relevancia el desarrollo de resultados enriquecidos para entornos de pregunta-respuesta (Enge, 2017). Los gestores de contenido se afanan en elaborar posibles resultados más relevantes y pertinentes que puedan ser seleccionados por el buscador. Podemos hablar del desarrollo de una nueva materia sobre la optimización de respuestas (*Answer Engine Optimization*) que explora los criterios de relevancia de los algoritmos de recuperación de información para proporcionar respuestas directas.

3. Metodología

Este trabajo ha acometido un estudio empírico de las respuestas enriquecidas en la SERP de Google centrado en analizar las respuestas directas y explorar el comportamiento del buscador para consultas en lenguaje natural del tipo “*What is?*”. Este análisis está basado en la observación y exploración de los resultados de la SERP de Google. Se examinan criterios como el tipo de respuesta, la disposición, el tamaño, la extracción total o parcial del párrafo, la ubicación del fragmento en el texto original, entre otras (véase la Tabla I). El procedimiento seguido para el análisis incluye las fases de: selección de consultas, recolección de datos y análisis de los resultados.

3.1. Selección de consultas

Se recopilaron 16 consultas de tipo informacional (Broder, 2002) en lenguaje natural y en inglés que incluyen las palabras “*What is?*”:

- Q1) What is NATO?
- Q2) What is Monkey Pox?
- Q3) What is Wordle?
- Q4) What is an Oligarch?
- Q5) What is an NFT?
- Q6) What is Aphasia?
- Q7) What is Hepatitis?
- Q8) What is the population of Ukraine?
- Q9) What is Salmonella?
- Q10) What is the Met Gala?
- Q11) What is a serp?
- Q12) What is a emotional breakdown?
- Q13) What is a degree in math?
- Q14) What is a handshake error?
- Q15) What is a ferromagnetic material?
- Q16) What is a career goal?

Se entiende que todas ellas tienen como propósito recuperar una respuesta específica. Para ello, se seleccionaron las preguntas de la lista de *Google Trends* del 2022 que se basan en términos de búsqueda cuyo mayor incremento de consulta se produjo en ese año en comparación con el anterior (Google, 2023; Brita-Paja, 2022). Las consultas contienen palabras clave simples y/o compuestas.

Bloque	Criterio	Descripción
SERP	Info + idiomas	Número de idiomas utilizados en los bloques
	Presencia <i>featured answer</i>	Presencia o no de una <i>featured answer</i> directa
	Imagen	Presencia de imagen junto a la <i>feature answer</i>
	1ª posición <i>feature answer</i>	Si la <i>feature answer</i> directa se encuentra en 1ª posición
	Posición <i>featured answer</i>	Posición de la <i>feature answer</i> en la disposición en la SERP
	Posición fuente en orgánicos	Posición del recurso fuente en los resultados orgánicos
	N_Bloques	Número de recursos enriquecidos
	Nº preguntas sugeridas	Nº de preguntas sugeridas por el buscador distintas
	Tarjeta de conocimiento	Fuente de la tarjeta (Wikipedia, Google)
	Imagen tarjeta	Presencia de imágenes
	Clasificación entidad	Clasificado como tipo de entidad
	GC	Gráfico de conocimiento
	Gráfico	Respuesta en forma de gráfico
	Calculadora	Si se muestra el <i>widget</i> de la calculadora
	Traductor	Si se muestra el <i>widget</i> para traducir términos a otros idiomas
	Diccionario	Si se muestra el <i>widget</i> para buscar términos
<i>Featured answer</i>	Polisemia <i>Keyword</i>	Si en los resultados se muestran varias acepciones
	Longitud	Número de palabras que contiene la respuesta
	Negrita	Si se utiliza formato negrita para presentar la respuesta
	Formato subrayado	Uso formato subrayado para presentar la respuesta
	Respuesta en origen	Nº de párrafos
Total/parcial		Respuesta seleccionada total o parcial del párrafo
Posición párrafo		Posición del párrafo respecto a la página fuente
Marcas en fuente		Marcas de formato del texto seleccionado en la página fuente
<i>Keywords</i> en encabezado		Presencia de la <i>keyword</i> en el título o encabezados
Negrita	Presencia de la <i>keyword</i> de consulta en formato negrita	

Tabla I. Criterios para análisis de respuestas directas

3.2. Recolección de datos

El estudio fue desarrollado durante el periodo comprendido entre los meses de febrero y marzo de 2023. Una vez ejecutadas las consultas se procedió a examinar los bloques de respuestas y se recopilaron los datos. Para cada respuesta del buscador se recogió información acerca de cómo se mostraban los resultados, si aparecían respuestas directas y contenido enriquecido, cuál era el contenido presentado, el recurso o recursos originales de las que proceden y la posición en el listado de la SERP.

3.3. Análisis de los resultados

El análisis de los datos se realizó mediante la observación y la exploración de características principalmente de los fragmentos destacados y su relación con otros elementos de respuestas enriquecidas a varios niveles (Tabla I, en la página anterior). Por un lado, los indicadores definidos para este estudio se dirigieron a observar la aparición de respuestas enriquecidas en la SERP, con respecto al tipo de contenido y cómo se presentaban en función de las palabras clave de la consulta. También se recogió información sobre el texto de la respuesta en la fuente de origen, así como la posición que ocupaba en el listado ordenado de los resultados orgánicos. Además de información presentada en los distintos GAB como imágenes gráficas y *widgets*, se anotó la posición y disposición que ocupaban.

Por otro lado, se han recogido factores de posicionamiento como la presencia de palabras clave de la consulta en encabezados, presencia de palabras clave en formato negrita, marcas de formato destacado y la posición del párrafo en el recurso. Para un examen más profundo de las respuestas directas se extrajo el texto y se analizó tamaño y la estructura lingüística (Sánchez-Cuadrado et al., 2003).

4. Resultados

A fin de mostrar los resultados obtenidos, se concretan a continuación los criterios en los que han sido clasificados de acuerdo con los objetivos: describir los tipos de respuestas enriquecidas, explorar los resultados, análisis de los factores de relevancia.

4.1. Tipo de respuestas enriquecidas y anatomía de la SERP

Entre las primeras posiciones de la página de resultados se incluyen respuestas directas alternativas como: un bloque de definición del término consultado, un bloque de tarjeta procedente de la Wikipedia, bloques y grafos de conocimiento, bloque

para la respuesta de la consulta o un bloque de preguntas (Figura 1). En cada disposición se incorporan respuestas enriquecidas y *widgets* que proporcionan funcionalidades concretas (calculadora, búsqueda de palabra en un diccionario, traducción de términos). A la hora de visualizar los resultados en la versión de escritorio, se observa un cambio en la anatomía de la SERP con respecto a versiones previas. Aunque, no todas las consultas devuelven respuestas destacadas (*featured answer*).

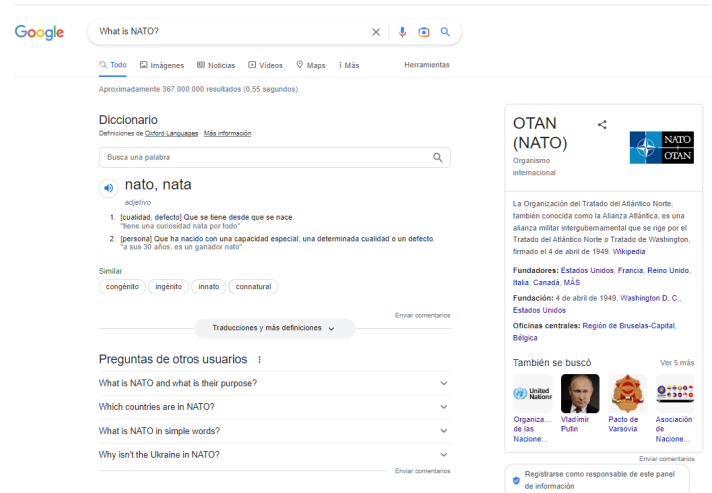


Figura 1. Resultado de consulta en Google para la pregunta 1

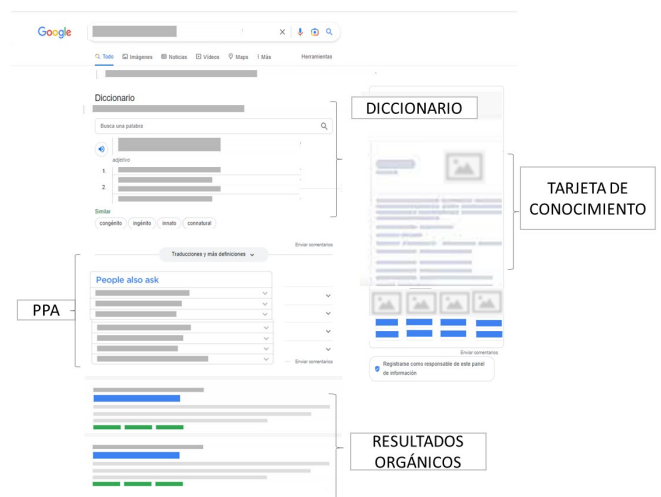


Figura 2. Anatomía de la SERP tipo 1

En cuanto a la ubicación de los resultados y las respuestas enriquecidas en la SERP se observan distintas disposiciones según las informaciones

que se vayan a ofrecer para cada consulta. Las respuestas con fragmentos enriquecidos suelen ocupar los primeros resultados, aunque también se observan algunos casos entre los resultados orgánicos comunes, y por tanto en posiciones intermedias. Las imágenes siguientes (Figuras 2, 3, 4 y 5) recogen cuatro ejemplos de disposición tipo que se han encontrado en la exploración de la SERP para los resultados a las consultas.

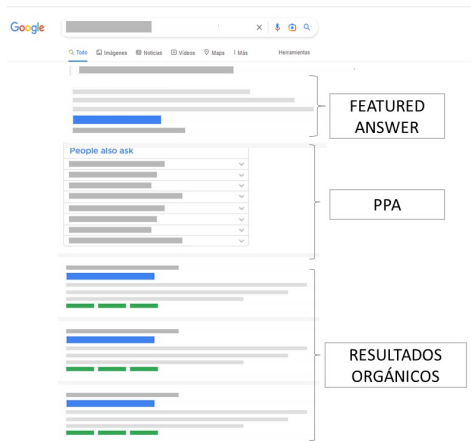


Figura 3. Anatomía de la SERP tipo 2



Figura 4. Anatomía de la SERP tipo 3



Figura 5. Anatomía de la SERP tipo 4

En la respuesta a la consulta sobre la población de un país se utiliza un gráfico dinámico para proporcionar un conocimiento complejo, basado en datos procedentes de diferentes fuentes (Figura 6). La incorporación de estos elementos implica que el listado de los resultados orgánicos queda relegado a la siguiente pantalla, teniendo que desplazar el *scroll* y cambiado de página para poder acceder a este tipo de resultados. Se observa como las respuestas directas mejoran las posibilidades de visualización en detrimento de los resultados posteriores.

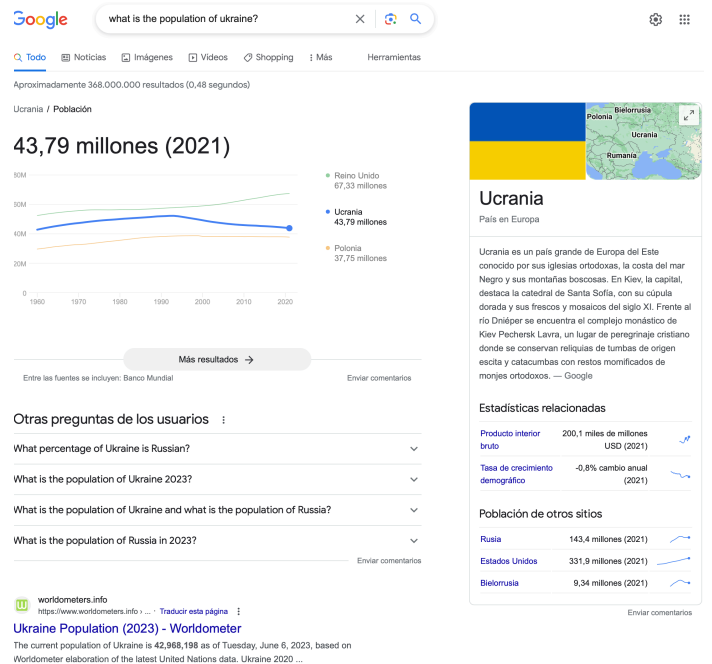


Figura 6. Resultado de la consulta 2 en Google

Explorando los resultados de las respuestas a las preguntas, se refleja control de la polisemia de las palabras. Se ha reportado el significado de un término común en el diccionario y en las tarjetas de la *Wikipedia* una entidad con nombre propio (por ejemplo, *OTAN/NATO*). Se observa además que el resultado puede ser plurilingüe basado con toda probabilidad en los datos de geolocalización de la ejecución de la consulta.

Se observa la concurrencia de recursos en idiomas diferentes en el 81 % de los casos, pudiendo mostrar los resultados orgánicos en el idioma de la consulta ejecutada (inglés) y tarjetas de información y grafos de conocimiento en otro. Por tanto, responde a procesos independientes del idioma con herramientas y mecanismos basados en el conocimiento que permiten identificar diferentes representaciones en un único concepto, así como personalizar los resultados por ubicación (Figura 7).

Google search results for "what is salmonella?". The search bar shows the query and the number of results (72,100,000). The results include a featured snippet from Wikipedia defining Salmonella as a bacterium, and several other links from Mayoclinic.org, CDC, and Clevelandclinic.org with snippets of text.

Figura 7. Resultado de la consulta 3 en Google

De todas las búsquedas se han recuperado respuestas enriquecidas de diversos tipos. En la mayoría de los casos se han identificado dos o tres bloques de respuesta enriquecida por consulta (Gráfico 1). En algo menos de la mitad, el buscador ha respondido con tarjetas de conocimiento procedentes del diccionario de Google y de la Wikipedia. Aunque solo en el 75 % de los casos, el buscador ha proporcionado un respuesta o fragmento destacado.

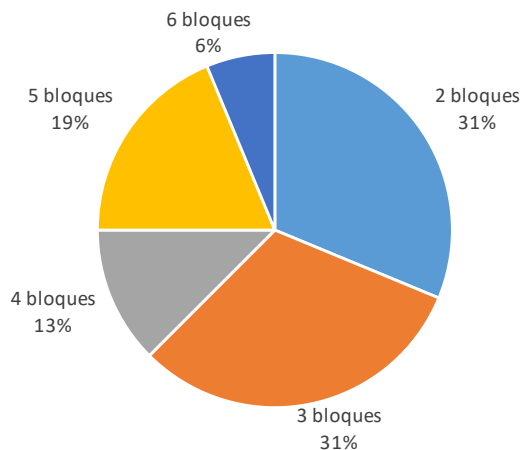


Gráfico 1. Proporción de bloques con respuestas enriquecidas por consulta

En todos los casos, el motor de búsqueda ha sugerido un bloque de PAA. Las preguntas sugeridas responden a los 5 tipos de preguntas (*what, why, who, how, when*), pudiendo mostrar más de 20 por bloque, aunque en la mayoría de los casos (43 %) se muestran 4. Cada una de estas preguntas puede ser respondida por distintas fuentes de información web.

También se detecta una correlación entre la estructura gramatical de la consulta y las otras preguntas de los usuarios que construyen frases con los mismos términos y estructuras semánticamente relacionadas. Se observa la generación de búsqueda amplia con todas las palabras de la consulta original (Tabla II) pero también se percibe un conocimiento semántico sobre el concepto. En la Tabla se muestra el resultado de la pregunta *What is a Salmonella?* Las preguntas sugeridas permiten deducir sin leer la respuesta que se trata de una cuestión relacionada con la comida y la salud. Se deduce además que tiene una sintomatología y una cura asociada. La configuración de la respuesta se expande a otros términos interrogativos frecuentes para proporcionar una visión facetada y multidimensional del concepto.

Consultas	Pregunta sugerida (PAA)
What is NATO?	What is NATO and what is their purpose? What countries are in NATO? What countries are not in NATO?
What is Wordle?	How do you play Wordle? What is Wordle and how does it work?
What is an NFT?	What exactly is an NFT in simple terms?
What is Salmonella?	What is the main cause of Salmonella? What happens if you get Salmonella? What food is Salmonella found in? What are 4 symptoms of Salmonella? What kills Salmonella in the body? Can Salmonella be cured?
What is the Met Gala?	What is the purpose of the Met Gala?

Tabla II. Ejemplos de preguntas sugeridas por Google para la consulta

4.2. Presentación de la featured answer

Para las respuestas destacadas, Google visualiza la respuesta directa en la SERP. La respuesta que define el término cuestionado se presenta en forma de párrafo. El GAB puede contener el párrafo de la respuesta, el nombre y el enlace al recurso fuente, información sobre la fuente, el título de la url seleccionada y una fecha, correspondiente a la fecha de actualización de la página. Se mantiene el contenido del párrafo original, pero se destaca la respuesta precisa en negrita, aunque en el recurso original no aparezca esa marca. El formato destacado puede referirse a todo el párrafo, o bien a una parte de este. En otras ocasiones, el texto completo de la respuesta se muestra en negrita, y se destaca con subrayado algunas palabras clave.

En cuanto al texto se identifica una relación entre el patrón sintáctico de las frases, de la pregunta

y las respuestas. Se recogen estructuras sintácticas que denotan una correspondencia léxico-semántica por las cuales se identifican las relaciones entre los términos (por ejemplo, *means* (X,Y), *ensures_that* (X,Y), *is_a* (X,Y)). Esas estructuras permiten identificar patrones que revelan respuestas a consultas simples que responden al interrogante “*What is?*”.

4.3. Recurso fuente de la respuesta: posición y formato

En el caso de los fragmentos destacados, los datos obtenidos indican que la respuesta suele proceder de los primeros recursos posicionado en el ranking de resultados de la SERP de Google. Como se observa en el Gráfico 2, el 46% se situó en la posición 0. Los resultados de este estudio muestran que puede estar situado en otra posición del ranking incluso alejada del top10.

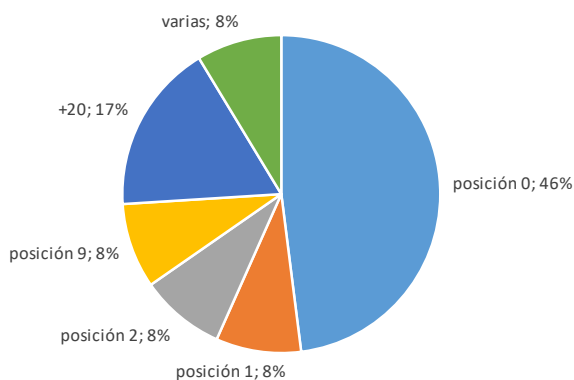


Gráfico 2. Posición del recurso fuente en el listado de resultados orgánicos

En relación con estrategias SEO *on-page*, tampoco se observa una relación directa con la posición estratégica del primer párrafo. Ni siquiera en el caso de las entradas de la Wikipedia usadas para dar respuesta a la pregunta se selecciona necesariamente el primer párrafo coincidente con la palabra clave de la consulta.

Para las respuestas destacadas se observa que la frase extraída puede estar ubicada en cualquier lugar de la página web, ya sea al inicio del párrafo, en una zona intermedia o al final. Se advierte que, en la mayoría de los casos examinados, la palabra clave de la consulta forma parte del encabezamiento del párrafo donde se encuentra la respuesta. Aunque el párrafo puede estar dentro de un bloque de contenido cuyo epígrafe no contiene los términos de la consulta, y la pregunta no es idéntica a la *query* ejecutada en el buscador. El código fuente de la página recoge el párrafo en un bloque de contenido `<div>` y contiene la palabra clave consultada en etiquetas

de encabezamiento `<h2>` y `<h4>`. En otras ocasiones aparece como un párrafo común y precedido de un encabezamiento que contiene parcialmente el término clave consultado en el caso de que sea compuesto. La longitud de la respuesta está entre 37-55 términos, aunque en ocasiones se ha observado un resultado menor (± 22 palabras) y no coincide necesariamente con un párrafo completo, pudiendo estar formada por partes de dos párrafos, o varias fuentes en el caso de los gráficos.

5. Conclusiones

Este estudio ha explorado cómo se presentan las respuestas directas para preguntas de carácter informacional del tipo *What is?* en dispositivos de escritorio. A lo largo de este trabajo se ha tratado de clarificar la tipología de los bloques de contenido para proporcionar respuestas directas (respuesta corta con texto, imágenes, *widgets* para la búsqueda del término en un diccionario, traductores, respuestas destacadas y otras preguntas de los usuarios). Con respecto a los resultados devueltos por Google, se han observado diversos tipos de *rich answers* para las preguntas *What is?* que pueden aparecer simultáneamente en la SERP para la misma consulta. Los datos indican que para todas las consultas se sugieren PAA. Estas preguntas sugeridas pueden estar situadas en posición inicial, intermedia y final, o incluso con varios bloques de preguntas a la vez. Estas preguntas tienen un carácter dinámico generando nuevas consultas relacionadas, lo que constituye un nuevo modo de navegar en la red.

A partir del análisis de la página de resultados, se ha demostrado la incorporación de múltiples bloques de respuestas directas. Por consulta se han recuperado de 1 a 5 bloques de respuestas en la primera página de resultados. Aunque lo más frecuente es que aparezcan entre 2 y 3 bloques de respuestas enriquecidas. La incorporación de estos contenidos en la SERP reduce considerablemente el espacio a los resultados orgánicos comunes, y en consecuencia se produce un desplazamiento del resto de resultados. Se ha observado que la Wikipedia no aparece como primer resultado, aunque aún mantiene una presencia significativa con las tarjetas de conocimiento laterales, y por tanto una alta visibilidad. Por lo tanto, entendemos que también experimenta este desplazamiento, manifestado por la pérdida de la primera posición en los resultados orgánicos listados.

Las respuestas directas correspondientes a las *featured answers* influyen de forma positiva en el buscador al proporcionar un resultado pertinente

directo con cero clics. Estas respuestas destacadas también ayudan a que el usuario tenga confianza del sitio, pues el sistema le otorga un espacio relevante en la SERP. A menudo este espacio se refuerza además con otras preguntas de los usuarios similares en un bloque contiguo o entre los resultados orgánicos. Con respecto a las *featured answers*, los datos indican que existe relación entre la selección como respuesta más relevante y posicionamiento SEO del recurso. Sin embargo, también parecen intervenir otros factores que alteran esa relación directa y que el fragmento provenga del primer resultado posicionado.

En la mayoría de los casos, la respuesta destacada (*featured answers*) procede del resultado que ocupa la primera o segunda posición. El recurso del que ha sido tomada la respuesta puede estar situado en cualquier posición de la primera página de resultados o incluso en páginas web posteriores. Por esto, las páginas que aplican estrategias SEO para lograr una posición entre los primeros resultados tienen una mayor probabilidad de ser visualizadas como una respuesta enriquecida. Sin embargo, parecen tener algunas peculiaridades. En consecuencia, se precisa profundizar en la optimización de preguntas en motores de búsqueda (AEO) y dilucidar cómo conseguir visibilidad de una página a través de respuestas destacadas.

Posicionar preguntas es una estrategia de posicionamiento muy acertada para el que consigue la posición 0, porque presenta una respuesta que resuelve de forma directa la necesidad del usuario, atrae tráfico web y puede mejorar la tasa de conversión de productos relacionados.

Las respuestas destacadas directas suponen un nuevo cambio en la búsqueda web. Con el análisis se advierte que la respuesta seleccionada no coincide con un párrafo completo. En ocasiones se seleccionan varios fragmentos de forma total o parcial, lo que inevitablemente plantea la cuestión de cuáles son los criterios para estimar que la respuesta está completa y como establecer los límites de la respuesta. Estos interrogantes evidencian que nos encontramos ante un sistema algorítmico más complejo.

Los avances en el procesamiento del lenguaje natural y la extracción de información están contribuyendo al desarrollo de criterios que determinen la relevancia informacional en los sistemas de recuperación de información. Los datos relativos al uso incremental de estrategias dirigidas a los asistentes de voz demuestran una clara apuesta de los motores de búsqueda por proporcionar respuestas directas. Los principales efectos que provoca son la mejora de la experiencia

del usuario y el suministro de respuestas a los asistentes de voz conversacionales.

En el caso de Google, las respuestas enriquecidas no solo proporcionan una valoración positiva sobre el buscador, sino que también se percibe una permanencia más duradera en la SERP. El usuario trata de explorar las distintas opciones ofrecidas por las *rich answers*, *widgets*, grafos de conocimiento y sugerencias de búsqueda, tratando de conceptualizar y contextualizar las consultas mediante relaciones y facetas esenciales.

Con la presente investigación se han obtenido resultados que confirman un cambio en la forma de determinar la relevancia de la información y de mostrar los resultados. Aunque el análisis de este estudio preliminar ha revelado datos representativos para establecer cambios en el algoritmo, entendemos que aumentar la muestra podría facilitar relaciones estadísticas más significativas. También sería de interés analizar las variables relativas a enlaces entrantes y dominio de autoridad que permitan observar el impacto de factores externos. Google hace cambios en el algoritmo y el aumento de nuevos recursos y preguntas puede alterar la disposición de los elementos en su página de resultados. Con todo, los resultados son válidos por cuanto muestran una situación en una determinada franja temporal y aunque haya modificaciones seguirá siendo una fase transitoria hacia otro modelo de recuperación en el que entrará en juego el posicionamiento de bloques de respuestas enriquecidas. Para futuros trabajos nos proponemos un estudio en profundidad sobre los criterios de selección de las respuestas destacadas y las otras preguntas de los usuarios. Otro tipo de búsquedas como *“How to?”* también merecen un análisis particular, pues presentan características diferenciadas en forma de listas para reflejar los pasos del procedimiento.

Referencias

- Bernstein, Michael S.; Teevan, Jaime; Dumais, Susan; Liebling, Daniel; Horvitz, Eric (2012). Direct answers for search queries in the long tail. // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Austin Texas USA: ACM 237-246. ISBN 9781450310154. <https://doi.org/10.1145/2207676.2207710>
- Bilal, Dania; Huang, Li-Min (2019). Readability and word complexity of SERPs snippets and web pages on children's search queries: Google vs Bing. // *Aslib Journal of Information Management*. 71:2, 241–259. <https://doi.org/10.1108/AJIM-05-2018-0124>
- Bink, Markus; Zimmerman, Steven; Elsweller, David (2022). Featured Snippets and their Influence on Users' Credibility Judgements. // *ACM SIGIR Conference on Human Information Interaction and Retrieval*, 113–122. <https://doi.org/10.1145/3498366.3505766>
- Brita-Paja Nuñez, P. (2022). Elementos de posicionamiento de los sistemas de pregunta respuesta de la web. *Getafe:*

- Universidad Carlos III de Madrid. Trabajo Fin de Grado. <http://hdl.handle.net/10016/36335>
- Broder, Andrei (2002). A taxonomy of web search. // ACM SIGIR Forum, 36:2, 3–10. <https://doi.org/10.1145/792550.792552>
- Chen, Wei-Fan; Hagen, Matthias; Stein, Benno; Potthast, Martin (2018) A User Study on Snippet Generation: Text Reuse vs. Paraphrases. // The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 1033–1036. <https://doi.org/10.1145/3209978.3210149>
- Duong, Véronique (2019). State of the Art of SEO. // SEO Management: Methods and Techniques to Achieve Success (1.a ed., pp. 1-9). Wiley Data and Cybersecurity. <https://doi.org/10.1002/9781119681427>
- eMarketer, & Insider Intelligence. (April 13, 2022). Number of voice assistant users in the United States from 2017 to 2022 (in millions) [Gráfico]. // Statista. <https://www-statista.com/statistics/1029573/us-voice-assistant-users/>
- Enge, Eric; Spencer, Stephan; Stricchiola, Jessie C. (2015). The Art of SEO: mastering search engine optimization. 3ª ed.. O'Reilly Media, Inc, USA.
- Enge, Eric (2017). Featured Snippets: New Insights, New Opportunities. Stone Temple Company Blog. <https://www.stonetemple.com/featured-snippets-new-insights-new-opportunities/>
- Epstein, Robert; Lee, Vivian; Mohr, Roger; Zankich, Vanessa R. (2022). The Answer Bot Effect (ABE): A powerful new form of influence made possible by intelligent personal assistants and search engines. // PLOS ONE. 17:6, e0268081. <https://doi.org/10.1371/journal.pone.0268081>
- Google (2023). Google Trend. <https://trends.google.es/>
- Harto, Agus Budi (2019). Implementing Website Design Based on Search Engine Optimization (SEO) Checklist to Increase Web Popularity. Journal of Applied Information, Communication and Technology, 6:2, 87–97. <https://doi.org/10.33555/ejaict.v6i2.67>
- Ishkin Rand. (2019 Agosto 13). Less than half of Google searches now result in a click. // SparkToro. <https://sparktoro.com/blog/less-than-half-of-google-searches-now-result-in-a-click/>
- Khshabi, Daniel; Ng, Amos; Khot, Tushar; Sabharwal, Ashish; Hajishirzi, Hannaneh; Callison-Burch, Chris (2021). GooAQ: Open Question Answering with Diverse Answer Types. // Findings of the Association for Computational Linguistics: EMNLP 2021, 421–433. <https://doi.org/10.18653/v1/2021.findings-emnlp.38>
- Killoran, John B. (2013). How to Use Search Engine Optimization Techniques to Increase Website Visibility. // IEEE Transactions on Professional Communication, 56:1, 50–66. <https://doi.org/10.1109/TPC.2012.2237255>
- Lewandowski, Dirk; Drechsler, Jessica; Mach, Sonja (2012). Deriving query intents from web search engine queries. // Journal of the American Society for Information Science and Technology, 63:9, 1773–1788. <https://doi.org/10.1002/asi.22706>
- Lurie, Emma; Mustafaraj, Eni (2018). Investigating the Effects of Google's Search Engine Result Page in Evaluating the Credibility of Online News Sources. // Proceedings of the 10th ACM Conference on Web Science, 107–116. <https://doi.org/10.1145/3201064.3201095>
- Mander, Jason; Buckle, Chase (2018). Voice Search. Insight Report 2018. // Global Web Index. <https://www.insidemarketing.eu/cdn/wp-content/uploads/2019/01/Voice-Search-global-web-indexreport.pdf>
- Miklošič, Andrej; Hlavatý, Ivan; Daňo, Ferdinand; Červenka, Peter (2016). Google Answer Box Keyword-related analysis. A case study. // European Journal of Science and Theology. 12:5,185-194. http://www.ejst.tuiasi.ro/Files/60/19_Miklosik%20et%20al.pdf
- Miklosik, Andrej. (2019). Search Engine Marketing Strategies: Google Answer Box-Related Search Visibility Factors. // Handbook of Research on Entrepreneurship and Marketing for Global Reach in the Digital Economy, edited by Luísa Cagica Carvalho y Pedro Isaías, Eds. Hershey, PA, USA: IGI Global, 2019. 463-485. <https://doi.org/10.4018/978-1-5225-6307-5.ch020>
- Mishra, Amit; Jain, Sanjay Kumar (2016). A survey on question answering systems with classification. // Journal of King Saud University - Computer and Information Sciences. 28:3, 345-361.
- Moldovan, Dan; Harabagiu, Sanda; Pasca, Marius; Mihalcea, Rada; Girju, Roxana; Goodrum, Richard; Rus, Vasile (2000). The structure and performance of an open-domain question answering system. // Proceedings of the 38th Annual Meeting on Association for Computational Linguistics - ACL '00, 563–570. <https://doi.org/10.3115/1075218.1075289>
- Morato, Jorge; Sánchez-Cuadrado, Sonia; Moreno, Valentin; Moreira, José. A. (2013). Evolución de los factores de posicionamiento web y adaptación de las herramientas de optimización. // Revista española de Documentación Científica. 36:3, e018. <https://doi.org/10.3989/redc.2013.3.956>
- Nayak, Pandu (2019). Understanding Searches Better Than Ever Before. // Google. <https://blog.google/products/search/search-language-understanding-bert>
- Nayak, Pandu (2022). How AI powers great research results. // Google. <https://blog.google/products/search/how-ai-powers-great-search-results/>
- Sánchez-Cuadrado, Sonia; Lloréns, Juan; Morato, Jorge; Hurtado, José. A. (2003). Extracción Automática de Relaciones Semánticas. 2da. Conferencia Iberoamericana // Sistemas, Cibernética e Informática. CISCi 2003. Orlando (Florida) 31 - 2 de agosto de 2003. 265-268.
- Singhal, Amit (2012). Introducing the knowledge graph: things, not strings. <https://googleblog.blogspot.co.at/2012/05/introducing-knowledge-graph-things-not.html>.
- StatCounter. (abril 26, 2023). Worldwide desktop market share of leading search engines from January 2015 to March 2023 [Gráfica]. // Statista. <https://www-statista.com/statistics/216573/worldwide-market-share-of-search-engines/>
- StatCounter. (mayo 10, 2022). Cuota de mercado de los principales motores de búsqueda online usados // España // 2021 [Gráfica]. // Statista. <https://es-statista.com/estadisticas/670092/cuota-de-mercado-de-los-motores-de-busqueda-por-buscador-espana/>
- Strzelecki, Artur; Rutecka, Paulina (2019). The Snippets Taxonomy in Web Search Engines. // M. Pańkowska & K. Sandkuhl (Eds.). Perspectives in Business Informatics Research. 365, 177-188. Springer International Publishing. https://doi.org/10.1007/978-3-030-31143-8_13
- Strzelecki, Artur; Rutecka, Paulina (2020a). Direct Answers in Google Search Results. // IEEE Access. 8, 103642-103654. <https://doi.org/10.1109/ACCESS.2020.2999160>
- Strzelecki, Artur; Rutecka, Paulina (2020b). Featured Snippets Results in Google Web Search: An Exploratory Study. In Á. Rocha, J. L. Reis, M. K. Peter, & Z. Bogdanović (Eds.), Marketing and Smart Technologies. 9-18. Springer. https://doi.org/10.1007/978-981-15-1564-4_2
- Sullivan, Danny (2018, enero 30). A reintroduction to Google's featured snippets. Blog Google. Consultado 30/03/2023. [Online] Disponible <https://blog.google/products/search/reintroduction-googles-featured-snippets/>

- Trippas, Johanne R.; Spina, Damiano; Thomas, Paul; Sanderson, Mark; Joho, Hideo; Cavedon, Lawrence (2020). Towards a model for spoken conversational search. // *Information Processing & Management*. 57:2, 102162. <https://doi.org/10.1016/j.ipm.2019.102162>
- Wu, Zhijing; Sanderson, Mark; Cambazoglu, B. Barla; Croft, W. Bruce; Scholer, Falk (2020). Providing Direct Answers in Search Results: A Study of User Behavior. // *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 1635-1644. <https://doi.org/10.1145/3340531.3412017>
- Yu, Liyang (2014). *A developer's guide to the semantic web (Second)*. Springer Berlin, Heidelberg. <https://doi.org/10.1007/978-3-662-43796-4>
- Zhao, Yiming; Zhang, Jin; Xia, Xue; Le, Taowen (2019). Evaluation of Google question-answering quality. // *Library Hi Tech*. 37:2, 308–324. <https://doi.org/10.1108/LHT-10-2017-0218>
- Ziakis, Christos; Vlachopoulou, Maro; Kyrkoudis, Theodosios; Karagkiozidou, Makrina (2019). Important factors for improving Google search rank. // *Future Internet*. 11:2, 32. <https://doi.org/10.3390/fi11020032>

Enviado: 2023-03-31. Segunda versión: 2023-05-26.
Aceptado: 2023-05-06.

Mención a la bibliografía en los metadatos de representación de objetos museables

Bibliographic notes in the representation metadata of museum objects

Beatriz TARRÉ ALONSO (1), Camila MONTEIRO DE BARROS (2)

Programa de Pos-graduación en Ciencia de la Información, Universidad Federal de Santa Catarina, R. Eng. Agrônomo Andrei Cristian Ferreira, s/n - Trindade, Florianópolis - SC, 88040-900, Brasil, bettytarrealonso@gmail.com. (2) camila.c.m.b@ufsc.br

Resumen

Las categorías de información sobre los objetos museables se representan en la web a través de metadatos, y es en la práctica de la documentación museística que estos aspectos aparecerán o no como facilitadores del acceso y comprensión del objeto por parte del usuario. Entre los diversos aspectos a describir sobre el objeto de museo (como autoría, material, iconografía, entre otros), el foco de esta investigación está en los metadatos de "bibliografía", ya que este tipo de referencia vincula el objeto con otra información, ampliando, profundizando o reafirmando su contexto original. El objetivo de esta investigación es verificar cuáles y cómo los museos de arte utilizan los metadatos "bibliográficos" en el conjunto de elementos descriptivos de los objetos museables que componen sus colecciones. Los resultados indican que, de un total de 34 museos analizados de diferentes partes del mundo, 14 incluyen bibliografía y 20 que no cuentan con ningún tipo de referencia. Además, 28 museos no contienen ningún otro vínculo con materiales documentales y solo 6 mantienen y consideran importante la inclusión de bibliografía complementaria.

Palabras clave: Descripción de obras de arte. Recuperación de información. Metadatos bibliográficos.

1. Introducción

A partir de la posibilidad de acceder a los objetos museables a través de imágenes accesibles en la web, la comunicación e información sobre estos objetos adquiere una importancia destacada en el pensamiento museológico. Teniendo en cuenta que en el museo el objeto adquiere el estatus de documento y soporte de información de variados rangos (Ceravolo y Tálamo, 2007), es en la documentación donde radica la oportunidad de registrar y hacer accesible ese rango de información.

Para Miranda (2022) la documentación museística se refiere al proceso de registro de objetos y colecciones que, a su vez, como afirman Ceravolo y Tálamo (2007), engloba una serie de procedimientos técnicos con fines de gestión, salvaguarda y disponibilidad de acceso a las colecciones de los museos. En este sentido, además de preservar el contexto original del objeto, se crea

Abstract

The categories of information about museum objects are represented on the web through metadata, and it is in the practice of museum documentation that these aspects will or won't appear as facilitators of access and understanding of the object by user. Among various aspects to describe about the museum object (such as authorship, material, iconography, and others), the focus of this research is on the "bibliography" metadata, since this type of reference links the object with other information, expanding, deepening or reaffirming its original context. The objective of this research is to verify which and how art museums use "bibliographic" metadata in the set of descriptive elements of the museum objects that make up their collections. The results indicate that of 34 museums analyzed from different parts of the world, 14 include a bibliography and 20 do not have any type of reference. In addition, 28 museums do not contain any other link with documentary materials and only 6 maintain and consider the inclusion of complementary bibliography important.

Keywords: Works of art description. Information retrieval. Bibliographic metadata.

un nuevo contexto (Navarrete y Owen, 2016), que puede representar la ampliación o profundización o reafirmación del contexto original, a partir de la conexión de este objeto con otros objetos (dentro y fuera de la colección del museo) y con otras categorías de información. Tales categorías circulan entre aspectos perceptibles y aspectos conceptuales interpretativos. No obstante, es en la práctica de la documentación que estos aspectos aparecerán o no como facilitadores del acceso y comprensión del objeto por parte del usuario. Para Ceravolo y Tálamo (2007), pensar en los museos también dirigidos al público externo es parte del propio proceso de musealización y está en el ámbito de la representación de la información sobre los objetos museables, con miras a su acceso y recuperación, aspectos analizados en la presente investigación.

2. Marco teórico

Las categorías de información sobre los objetos museables se representan en la web a través de metadatos. Desde la perspectiva semántica susceptible de ser aprovechada en la web enlazando información de diferente tipo, se modificó la percepción del documento. Es posible, a través de la representación de metadatos, crear un cierto tipo de discurso sobre el objeto. Según Miranda (2022, p. 4), el conjunto de metadatos, como uno de los resultados de la documentación museística, sirve para “registrar discursivamente a maior quantidade de informações contidas em um determinado bem cultural, incluindo elementos relacionados ao seu significado e ao seu significante.” Es decir, son los metadatos los que promoverán la base contextual del objeto, y el usuario necesita esta descripción tanto para poder recuperar la información deseada como para entender el objeto de una forma más amplia y plural (Navarrete y Owen, 2016).

Sin embargo, Silva y Lara (2021) llaman la atención sobre la falta de bases metodológicas en los museos, especialmente brasileños, para la práctica de la descripción, encaminada a brindar servicios dirigidos al público externo y también al público interno del museo, como curadores, museólogos, gestores, entre otros. Existen algunos documentos que brindan las bases para la mejor elección de los metadatos de representación tanto en términos de la información a describir como de los estándares sobre cómo describir cada pieza de información. Las autoras compararon tres esquemas de metadatos destinados a museos: SPECTRUM, que es un estándar de metadatos utilizado por *Collections Trust*, en el Reino Unido; las categorías de información del Comité Internacional de Documentación (CIDOC-ICOM); y *Categories for the Description of Works of Art* (CDWA) por el *Getty Research Institute*.

Entre varios metadatos comparados en cuanto a la ocurrencia y similitud entre los patrones analizados por Silva y Lara (2021), llama la atención la presencia de los metadatos “bibliografía” en los tres casos para la presente investigación. Además de la importancia de registrar las fuentes que sustentan la descripción del objeto, también se trata de indicar otra información relacionada que pueda enriquecer cualitativamente la interpretación del objeto por parte del usuario. Dichas referencias pueden ser artículos y libros publicados o inéditos, manuscritos, imágenes, audios, catálogos de exposiciones, entre otros materiales.

Según Navarrete y Owen (2016, p. 115), “*it is the metadata attributed to the objects that enables discoverability via crossreferences, hyperlinks,*

multiple interpretations [...]” y también la relevancia de que el museo brinde apoyo informativo a quienes acceden a su colección en la web, el objetivo de esta investigación es verificar cuáles y cómo los museos de arte utilizan estos metadatos en el conjunto de elementos descriptivos de los objetos museables.

3. Metodología

Esta investigación tiene un enfoque cuantitativo, ya que permite medir, relacionar, verificar y predecir la naturaleza de la representación de los objetos museables como elementos informativos. Tiene un carácter descriptivo, ya que busca especificar las propiedades, características y dimensiones de un fenómeno de representación de información. Se centra en el análisis de los sistemas de representación y recuperación de colecciones de arte, actualmente utilizados por museos y galerías. Tiene un diseño no experimental, debido a que las variables no son manipuladas deliberadamente. El corte transversal se realiza recogiendo datos en un único período de tiempo y analizando su incidencia en el momento estudiado.

Para examinar la representación bibliográfica de los objetos museables, se seleccionó un conjunto de museos y galerías. Para justificar su selección, se revisó el sitio web del Consejo Internacional de Museos (ICOM), en busca de un ranking de los principales museos de arte o un directorio internacional que permitiera acceder a datos como el número de visitantes. Como la estrategia anterior no fue efectiva, se realizó una búsqueda en *Google Arts & Culture*, considerado el proyecto que aglutina e incluye contenidos de más de 2.000 museos, instituciones y galerías. La muestra seleccionada fue tomada como referencia en otras investigaciones realizadas (Tarré, Díaz-Redondo y Monteiro, s. f., enviado para publicación).

Las instituciones identificadas se encuentran en la Tabla I, en la página siguiente, que ilustra la distribución por países. La muestra fue seleccionada aleatoriamente y está compuesta por 34 museos y galerías, pero tuvo en cuenta la representación geográfica de Europa, América del Norte y el Caribe. Para la propia viabilidad del presente estudio se limitó la exhaustividad de la muestra seleccionada, y el análisis comparativo de América Central y del Sur está previsto para futuras investigaciones, el cual garantizará ampliar el foco de la investigación.

Los datos fueron recolectados del sitio web oficial de cada institución. Se realizaron búsquedas de varias obras y los resultados de la recolección fueron tabulados.

<i>País</i>	<i>Museo</i>
Alemania	Museo Städel
Austria	Museo de Historia de Arte de Viena
Bélgica	Museo del Cincuentenario de Bruselas
España	Museo del Prado
	Museo Nacional Centro de Arte Reina Sofía
	Museo Thyssen-Bornemisza
	Museo del Romanticismo
	Museo Guggenheim Bilbao
Cuba	Museo Nacional de Bellas Artes
Estados Unidos	MoMA
	<i>The MET - Metropolitan Museum</i>
	<i>Art Institute of Chicago</i>
	<i>Museum of Fine Arts of Boston</i>
	<i>National Gallery of Washington</i>
	<i>Denver Art Museum</i>
	<i>Philadelphia Museum of Art</i>
	<i>Detroit Institute of Arts</i>
	<i>Smithsonian American Art Museum</i>
Francia	Museo do Louvre
	Museo de Orsay
	Centro Georges Pompidou
Grecia	Museo de la Acrópolis
Italia	Galería Borghese
	Museo de Capodimonte
	Galería Uffizi
	Museos del Vaticano
Inglaterra	<i>British Museum</i>
	<i>The National Gallery</i>
	<i>Victoria and Albert Museum</i>
	<i>Tate Modern</i>
Países Bajos	<i>Rijksmuseum</i>
Portugal	Museo Serralves
	Museo Calouste Gulbenkian
Rusia	Museo Hermitage

Tabla I. Distribución de la muestra

4. Resultados

La muestra distribuida en Europa (24), América del Norte (9) y el Caribe (1), cuenta con un total de 13 nacionalidades diferentes. El Gráfico I presenta la naturaleza de las instituciones.

La variedad en la tipología de las colecciones se indica en el Gráfico II.

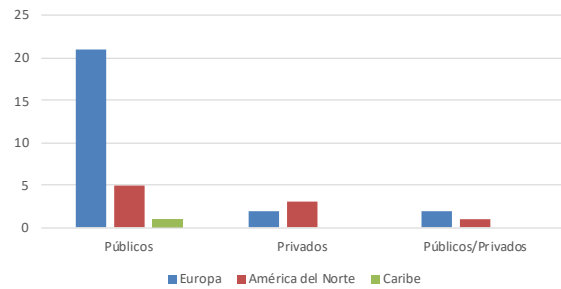


Gráfico I. Clasificación por regiones de la muestra de museos y galerías seleccionada

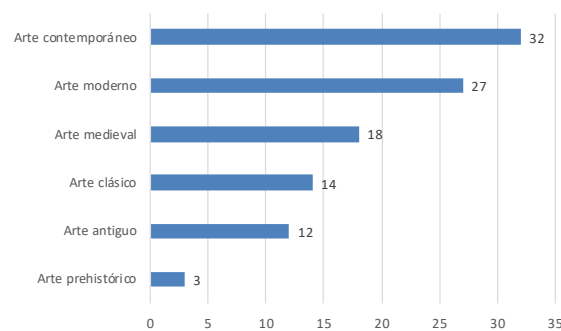


Gráfico II. Tipología de la muestra de museos y galerías seleccionada

Se observa un predominio de los museos que valoran el arte contemporáneo, frente al bajo predominio de los que tienen arte prehistórico. Por otro lado, de la muestra escogida, 3 museos cuentan con colecciones que van desde el Arte prehistórico hasta el Arte contemporáneo, que son los siguientes: Museo del Cincuentenario de Bruselas (Bélgica), Museo Nacional de Bellas Artes (Cuba) y Museo del Louvre (Francia). De las instituciones analizadas, 6 están dedicadas a los 3 últimos periodos (Arte medieval, Arte moderno y Arte contemporáneo); 6 a Arte moderno y Arte contemporáneo; 7 de ellos atesoran exclusivamente Arte contemporáneo; solo uno está dedicado al Arte clásico (Museo de la Acrópolis, Grecia) y uno al Arte moderno (Museo del Romanticismo, España).

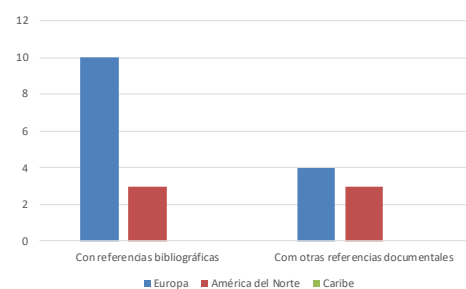


Gráfico III. Distribución de museos y galerías por región con referencias bibliográficas

Se analizó si dentro de los apartados dedicados a la descripción de las colecciones se incluyen referencias bibliográficas sobre la obra o cualquier otro material bibliográfico. Los resultados se muestran en el Gráfico III a continuación.

En la muestra estudiada se observa un predominio de instituciones europeas (10) con inclusión de referencias bibliográficas, sobre las norteamericanas (4). Sin embargo, con respecto a las otras referencias tanto Europa como América del Norte se encuentran igualadas (3, respectivamente); no encontrando ocurrencia de ninguno de los aspectos anteriores en el Caribe.

5. Discusión y conclusiones

El siguiente análisis se basa en criterios preestablecidos y está organizado por región geográfica, siendo Europa el primer lugar a abordar, por ser el área con mayor prevalencia de ocurrencia de los campos estudiados. Más detallada y concretamente sobre los museos y galerías que incluyen referencias bibliográficas se encuentra el Museo del Louvre (París, Francia) con la sección bibliográfica (pero sin enlaces) y, notas manuscritas con imágenes y archivos en formato pdf, con información que se puede descargar. El Museo de la Acrópolis (Atenas, Grecia) contiene bibliografía, pero tampoco presenta enlaces. El *British Museum* (Londres, Inglaterra) proporciona Referencias bibliográficas (*Bibliographic references*) con un enlace a un archivo de catálogo con metadatos en el mismo sitio web del museo. El Museo Calouste Gulbenkian, (Lisboa, Portugal) también incluye referencias, pero bajo el tradicional título de *Bibliografía*.

La *Galeria degli Uffizi* (Florencia, Italia) bajo el título *Hipervisiones* presenta un proyecto de narración (doce obras maestras de los Uffizi, narradas en bandas sonoras) y otros proyectos sobre pinturas y descripciones literales de la curaduría, mas sin referencias. También incluye videos explicativos de los curadores o artistas frente a las obras.

En España, el Museo Guggenheim Bilbao contiene notas con referencias, pero sin enlace, como muchas de las instituciones anteriores analizadas. En Madrid, el Museo del Prado presenta, inclusive, una cronología en su sitio web y contiene una bibliografía sobre las obras, aunque también carece de enlaces externos. El Museo Thyssen-Bornemisza permite descargar un archivo con el catálogo de la obra, pero sin las referencias bibliográficas. El Museo del Romanticismo dispone de un enlace a CER.ES, la *Red Digital de Colecciones de Museos de España*, que incluye la descripción de las colecciones,

pero manifiesta errores de accesibilidad. No contiene otro material bibliográfico, pero incluye un enlace a la red de imágenes de Flickr, que utiliza etiquetas como palabras clave y una breve descripción de la obra, mas sin referencias documentales.

El Museo del Cincuentenario de Bruselas (Bruselas, Bélgica) tiene el inconveniente de que su sitio web aún está en construcción y la parte dedicada a la descripción de las colecciones presenta solo algunos aspectos destacados de las obras (*highlights*), sin referencias bibliográficas. A esto se suma el hecho de que solo tiene un enlace al catálogo de obras en línea del museo sin tener ningún vínculo, por temática o de cualquier tipo, siendo demasiado general.

La *Tate Modern* (Londres, Inglaterra) a pesar de no incluir bibliografía, tiene una sección denominada Lectura adicional (*Further Reading*), pero sin enlaces para acceder a las referencias recopiladas. Por su parte, el Museo Städel (Frankfurt, Alemania) no presenta bibliografía ni información complementaria adicional, pero sí contiene metadatos y palabras clave que se subdividen en las categorías de motivos (*motifs*), referencia (*reference*), y también aplica el *iconclass* por niveles en la descripción.

El *Rijksmuseum* (Ámsterdam, Países Bajos) tiene una sección de Documentación (*Documentation*) que brinda acceso gratuito a la biblioteca *Rijksmuseum Research Library* (<https://library.rijksmuseum.nl/cgi-bin/koha/opac-main.pl>), que es un catálogo en línea que utiliza el formato interoperable MARC 21, y según la propia web, una de las bibliotecas de arte más importantes del mundo. Incluye el acceso a catálogos de subastas y exposiciones, catálogos comerciales y de colecciones, así como libros, periódicos y memorias relativas a las colecciones del museo, y se recopilan ininterrumpidamente desde 1885. El OPAC permite, entre sus funcionalidades, como muchas otras, guardar registros, comprar, verificar la disponibilidad en la biblioteca y comprobar el código de barras de las copias de documentos. El enlace, a diferencia de algunos de los anteriores, hace referencia específica a las bibliografías sobre el trabajo investigado. Además, contiene un clip de audio, con sonidos alusivos a las escenas dentro de la propia obra y que forma parte del Tour Multimedia (*Multimedia Tour*).

Por último, entre las instituciones europeas destaca la Galería Borghese (Roma, Italia), que contiene una bibliografía sobre su obra y algunas de ellas con enlaces. La procedencia incluye referencias a inventarios con datos, títulos, números y páginas, aunque con algunos textos solo en el idioma original (italiano), y enlaces a archivos de

Internet con los documentos digitalizados. Por otra parte, la Galería Borguese tiene una asociación en el proyecto *Internet Archive* (<https://archive.org/about/>), que es una iniciativa que comenzó en 1996. Como se define en su sitio web, es una biblioteca digital de sitios de Internet y otros artefactos culturales, en formato digital. Como biblioteca física, brinda acceso gratuito a investigadores, historiadores, académicos, personas con discapacidad y público en general; con la misión de proporcionar acceso universal a todos los conocimientos. Entre sus patrocinadores se encuentran asociaciones relevantes en el área de la Biblioteconomía como la *American Library Association* (ALA), la *International Federation of Library Associations and Institutions* (IFLA), y la *Music Library Association* (MLA), entre muchas otras. El proyecto permite interactuar con los documentos físicos que fueron digitalizados con una alta calidad y nivel de precisión y nitidez. Contiene el URN (*external-identifier*) como el número del documento en el catálogo mundial (OCLC), el registro (*full catalog record*) del formato MARCXML, además de otros identificadores y metadatos relevantes, que sugieren artículos similares basados en ellos. Además, permite escuchar la versión del audiolibro (*audio book*) e interactuar con él digitalmente. Entre las funcionalidades que brinda, cumpliendo con las licencias de derechos de autor, se encuentran: compartir en redes sociales, descargas en una amplia variedad de formatos, desde pdf hasta epub y muchos más. También presenta estadísticas, como el número de vistas realizadas en el documento, selección de favoritos, una vez iniciada la sesión en la plataforma, el intercambio de revisiones realizadas, interacción a través de comentarios, e incluso ofrece la posibilidad de ser parte del proyecto y participar en un área de interés.

En América del Norte, el *Art Institute of Chicago* (Chicago, Estados Unidos) presenta Historia de la Publicación (*Publication history*) e Historia de la exposición (*Exhibition history*) en las cabecezas, y como sus nombres lo indican, incluye una recopilación de las publicaciones donde se hace referencia a la obra y dónde fue expuesta, es decir, los catálogos de las exposiciones; pero es válido señalar que cada una de las entradas, actuales o no, carecen de la inclusión de enlaces internos o externos a un catálogo en línea. Por otra parte, la misma institución contiene una sección denominada Recursos educativos (*Educational resources*) con enlaces a archivos en formato pdf, que incluyen datos recopilados como: un resumen de la obra, la biografía del artista, un glosario, una serie de aplicaciones para la sala de aulas (*classroom applications*), otros conceptos interesantes y la bibliografía.

En Washington (Estados Unidos), la *National Gallery of Art*, en la sección de Bibliografía (*Bibliography*), incluye publicaciones con referencias a la obra y catálogos de exposiciones bajo el mismo título, a diferencia del *Art Institute of Chicago*. Además, permite la citación directa con un botón predeterminado, que ofrece una referencia automatizada; y la exportación de la información del sitio web a formato pdf, con la posibilidad de citarlo y compartirlo en redes sociales u otras plataformas. Adicionalmente, en el campo de Procedencia (*Provenance*), muestra notas de cartas, índices, artículos de prensa y catálogos relacionados con las obras o que apoyan el estudio de las colecciones, pero la mayoría sin enlaces para acceder a ellas. Por otro lado, las secciones dentro del sitio web de la biblioteca, archivo y centro de estudios, aunque pueden ofrecer un link a estos documentos digitalizados, se encuentran separadas, es decir, sin referencias cruzadas. También en Washington, el *Smithsonian American Art Museum* utiliza únicamente el método *Linked Open Data*, con la implementación del modelo conceptual de museos CIDOC.

El *Detroit Institute of Arts* (Michigan, Estados Unidos) tiene la particularidad de incluir bibliografía bajo el título de Referencia publicada (*Published reference*), pero más allá de eso no contiene materiales bibliográficos adicionales. Sin embargo, *The MET —Metropolitan Museum of Art—* (Nueva York, Estados Unidos) tiene en la sección Línea del Tiempo de Historia del Arte (*Timeline of Art History*), ensayos relacionados (*related essays*), con enlaces. Además de encabezados como citas (*citation*), lecturas adicionales (*further reading*), con enlaces y cronologías con referencias. También incluye una referencia cruzada a Publicaciones del Museo (*Museum Publications*) con notificaciones sobre títulos impresos (*This title is in print*) o en línea (*Read online*), que también permite la descarga en formato pdf y la consulta en catálogos en línea como *WATSONLINE* y *WorldCat*.

Entre las instituciones analizadas, los mejores ejemplos de representación documental bajo la inclusión de las variables consideradas son la *National Gallery of Art*, de Estados Unidos, el *Rijksmuseum*, de Holanda y la Galería Borguese, de Italia; dos de ellos europeos y solo uno norteamericano. Existe un número significativo que no incluye referencias bibliográficas y la diferencia es notoria, pues de un total de 34 en la muestra seleccionada, la prevalencia de criterios es de un 41,18 % (14 que incluyen), y un 58,82 % (20 instituciones) que no tienen ningún tipo de referencia. Sin mencionar que un 82,35 % (28) no contienen ningún otro vínculo con materiales docu-

mentales y, solo un 17,65 % (6) mantienen y consideran importante la inclusión de bibliografía complementaria. Lo anterior puede deberse a la desvinculación de la práctica museológica con la bibliotecológica, perjudicando directamente a la sinergia que produce la unión de las mismas, pues garantizaría una mejora evidente en la recuperación de información sobre obras de arte y contribuiría a los procesos investigativos y curatoriales de colecciones artísticas.

Referencias

- Baca, Murtha; Harpring, Patricia (eds.). *Categories for the Description of Works of Art* (2022). Los Angeles (EUA): Getty Vocabulary Program. https://www.getty.edu/research/publications/electronic_publications/cdwa/ (2023-03-25).
- Ceravolo, Suely Moraes; Tálamo, Maria de Fátima (2007). Os museus e a representação do conhecimento: uma retrospectiva sobre a documentação em museus e o processamento da informação. // *Encontro Nacional de Pesquisa em Ciência da Informação*, 8, Salvador, 28 a 31 de octubre de 2007. *Anales Electrónicos...* Salvador: UFBA. <http://professor.ufop.br/sites/default/files/mas/files/ceravolotalamo.pdf> (2023-03-25).
- CIDOC-ICOM (c2023). *International guidelines for museum object information: the CIDOC Information Categories*. Paris: International Committee for Documentation of the International Council of Museums. <https://cidoc.mini.icom.museum/> (2023-03-25).
- Collections Trust (2022). *Spectrum 5.0: the UK Museum collections management standard*. [Londres]: Collections Trust. <https://collectionstrust.org.uk/spectrum/spectrum-5/> (2023-03-25).
- Miranda, Rose Moreira de (2022). Quem procura acha?. // *Reflexões sobre a semântica na documentação museal*. 42 (2022). <https://revistas.unilasalle.edu.br/index.php/Mouseion/article/view/10059>. (2023-03-25).
- Navarrete, Trilce; Owen, John Mackenzie (2016). *The Museum as Information Space: Metadata and Documentation*. // Borowiecki, Karol Jan; Forbes, Neil; Fresa, Antonella (eds.). *Cultural Heritage in a Changing World*. Cham (Suiza): Springer International Publishing. 111-123. <https://library.oapen.org/bitstream/handle/20.500.12657/28109/1001885.pdf?sequence=1#page=135> (2023-03-25).
- Silva, Camila Aparecida da; Lara, Marilda Lopes Ginez de (2021). Esquema básico de metadados para representação descritiva de obras de arte em museus brasileiros. // *Transinformação*. 33 (2021). <https://doi.org/10.1590/2318-0889202133e200050> (2023-03-25).
- Collections Trust (c2023). *Spectrum*. <https://collectionstrust.org.uk/spectrum/> (2023-03-25).
- Tarré Alonso, Beatriz; Díaz-Redondo, Carlos; Monteiro de Barros, Camila [s.f.]. *Análise da influência da atribuição de metadados semânticos para a representação de informação sobre coleções de arte* [Enviado para publicación].

Enviado: 2023-03-29. Segunda versión: 2021-05-10.
Aceptado: 2023-06-06.

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.

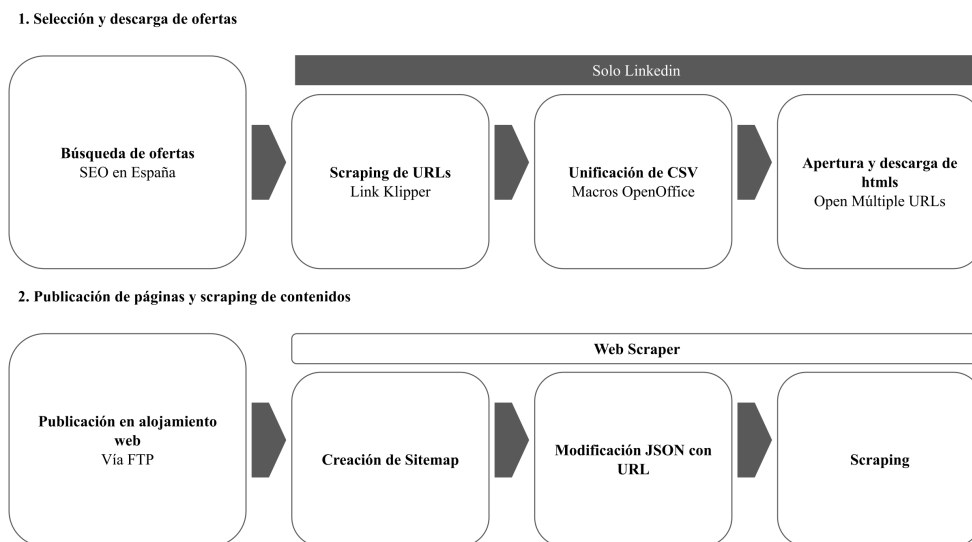


Figura 1. Procedimiento de descarga y scraping de ofertas laborales

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).

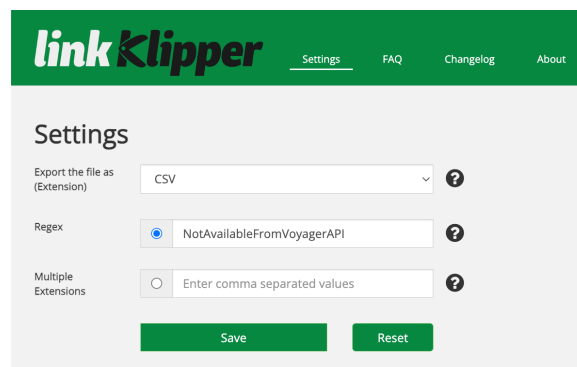


Figura 2. Panel de configuración de la herramienta Link Klipper (Link Klipper).

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. 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).

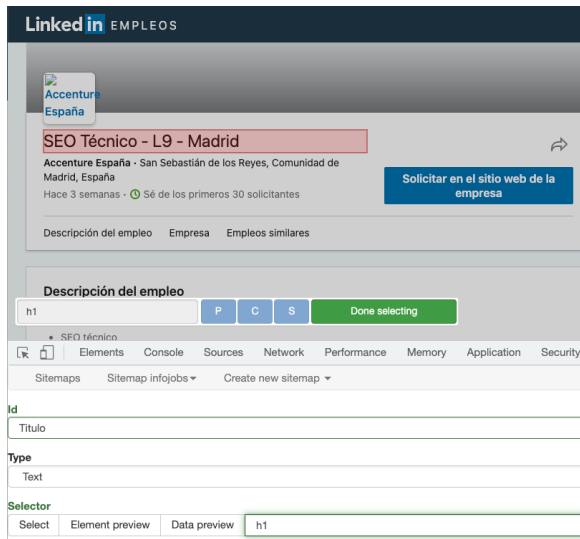


Figura 5. Captura del elemento título de una de las ofertas de trabajo en la herramienta Web Scraper (LinkedIn y Web Scraper)

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	Element preview
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	Element preview
ubicacion	div.col-6:nth-of-type(1) div.col-child > ul.list-default > li:nth-of-type(1)	SelectorText	no	_root	Element preview
fecha-publicacion	li:nth-of-type(3) span:nth-of-type(2)	SelectorText	no	_root	Element preview
descripcion	div.inner.panel-canvas	SelectorText	no	_root	Element preview
nivel-experiencia	span#prefijoNivelLaboral.list-default-text	SelectorText	no	_root	Element preview
sector-tipo-industria	div.inner.panel-canvas li:nth-of-type(1) span.list-default-text	SelectorText	no	_root	Element preview
tipo-empleo-contrato	span#prefijoJornada	SelectorText	no	_root	Element preview
funciones-categoria	div.inner.panel-canvas li:nth-of-type(2) span.list-default-text	SelectorText	no	_root	Element preview
salario	div.col-6 li:nth-of-type(1) > span	SelectorText	no	_root	Element preview
experiencia-minima	div.col-6:nth-of-type(2) li:nth-of-type(2) span	SelectorText	no	_root	Element preview
estudios-minimos	span#prefijoEstMin.list-default-text	SelectorText	no	_root	Element preview
conocimiento1	li:nth-of-type(1) a.tag	SelectorText	no	_root	Element preview
conocimiento2	li:nth-of-type(2) a.tag	SelectorText	no	_root	Element preview

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 (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 (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 (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.ijirccce.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 (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; Pastuszak, 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.

Crítica diplomática: um estudo panorâmico da produção científica

Crítica diplomática: un estudio panorámico de la producción científica

Diplomatic criticism: a panoramic study of scientific production

Sonia Oliveira Matos MOUTINHO, Sonia Maria TROITIÑO RODRIGUEZ

Universidade Estadual Paulista –Unesp, Campus de Marília, Dep. de Ciência da Informação, Av. Hygino Muzzi Filho, 737, Campus Universitário, 17525-900 –Marília, SP (Brasil). [sonia.matos-moutinho | sonia.troitino] @unesp.br

Resumen

Se aborda la crítica diplomática a partir de la producción científica publicada en bases de datos científicas internacionales. Para este estudio empleamos el método de la bibliometría aplicada para identificar la producción científica indizada acerca del objeto investigado. La investigación es exploratoria-descriptiva y cuali-cuantitativa. Como fuentes de investigación se utilizaron documentos recuperados de Google Scholar, Web of Science y Scopus por medio del software Publish or Perish, con el fin de verificar la producción teórica de la crítica diplomática y las métricas de interés. RawsGraph, un marco de código abierto utilizado para la visualización de datos complejos, se aplicó para la visualización de las métricas. Con ello se pudo comprobar que los estudios sobre Crítica Diplomática, en el contexto internacional, se encuentran en crecimiento exponencial y tienen sus primeros registros de publicación en el año 1681.

Palabras clave: Crítica diplomática. Producción científica. Estudios bibliométricos.

1. Introdução

A origem da Diplomática está ligada à necessidade de verificação de autenticidade de documentos mediante dúvidas e questionamentos sobre a veracidade dos registros de diplomas do medievo. De acordo com Bellotto (2002, p. 15), isso se deu quando jesuítas franceses, em 1643, liderados por Jean Bolland, resolveram publicar uma gigantesca história dos santos, a *Acta Sanctorum*, na qual se queria avaliar criteriosamente a vida dos santos a fim de separar a realidade das lendas.

Em 1675, quando escreveu a introdução do segundo volume do *Acta Sanctorum*, coleção destinada a analisar a biografia de santos católicos e separar evidências reais de lendas persistentes, o jesuíta bolandista Daniel Van Papenbroeck questionou a veracidade documental de diplomas pertencentes aos mosteiros beneditinos da Abadia de Saint Denis. Assim, foi estabelecida a crítica documental, na qual os documentos seriam tanto menos dignos de fé quanto mais antigos fossem, e pôs em dúvida algumas peças da

Abstract

Diplomatic Criticism is addressed based on scientific production published in international scientific databases. For this study, we used the method of applied bibliometry to identify the undisclosed scientific production on the investigated object. The investigation is exploratory-descriptive and qualitative-quantitative. As research sources, documents retrieved from Google Scholar, Web of Science and Scopus by means of the Publish or Perish software were used, in order to verify the theoretical production of the Diplomatic Criticism and the metrics of interest. RawsGraph, an open source framework used for visualizing complex data, is applied for visualizing metrics. With it, it was possible to confirm that studies on Diplomatic Criticism, in an international context, were experiencing exponential growth and had their first publication records in the year 1681.

Keywords: Diplomatic criticism. Scientific production. Bibliometric studies.

época merovíngia existentes nos arquivos da Ordem de São Bento.

Em resposta ao questionamento da autenticidade de documentos resguardados por mosteiros beneditinos, um dos membros de sua ordem, Jean de Mabillon, historiador e monge beneditino da congregação de Saint-Maur, de Paris, que se encontrava em Saint-Germain-des-Près justamente para publicar a vida dos santos de sua Ordem, publicou seis anos mais tarde, em 1681, um tratado em resposta àquela desconfiança. Trata-se de uma obra em seis partes intitulada *De re diplomatica libri VI*, que estabeleceu as regras fundamentais da crítica textual e serviu de berço para o surgimento de ciências tais como a Diplomática e a Paleografia. Mabillon, ao pesquisar em muitos arquivos eclesiásticos franceses, alemães e italianos, com seu trabalho metódico e objetivo, desenvolveu muitos dos procedimentos e técnicas destinados à comprovação de autenticidade, análise e compreensão dos atos escritos que, aperfeiçoados, foram sendo adota-

dos através dos tempos por outras ciências (Tallafigo, 1994; Duranti, 1995; Tessier, 1952; Valente, 1978). Nascia assim, muito imbricada em objetivos forenses e no Direito Eclesiástico, a Diplomática.

Desde então, a Diplomática tem evoluído em direção ao Direito, à Heurística e à Arquivística, modificando, para tanto, seus instrumentos e apresentando maior ampliação de objeto e campo de ação do que em sua origem. Tal ciência vem evoluindo há três séculos e meio, e que, devido às mudanças nas práticas de confecção e documentos diante das tecnologias digitais, coloca-se de suma importância um panorama da produção científica sobre a temática a fim de motivar a reflexão dos profissionais e pesquisadores da Arquivologia, da Paleografia e da Ciência da Informação.

A Diplomática do documento contemporâneo, segundo Tallafigo (1994, p. 21), não tem por que ser reservada apenas a historiadores da idade média ou da modernidade. Para ele (Tallafigo, 1994, p. 21):

Evidentemente por ahí, por el medievalismo se há consolidado y afilado sus armas como ciência em el arte de discernir y sistematizar los caracteres internos y externos de los documentos, siendo sus frutos impresionantes en cuanto a los conocimientos de instituciones y publicación de colecciones diplomáticas. Esas armas bien utilizadas y adaptadas a los documentos contemporáneos pueden hacer una historia al hilo de las formas documentales, como espejos de las sociedades y sus hombres em la contemporaneidad.

Portanto, mesmo tendo suas origens na Idade Média, os estudos sobre crítica diplomática são contemporâneos e possuem um longo caminho de estudos e pesquisas.

Com este trabalho busca-se contribuir para o avanço do conhecimento da própria temática Diplomática, objetivando propor indicadores, reflexões e análises relativos à produção intelectual sobre o tema. Ademais, partindo do questionamento sobre como os estudos sobre a crítica diplomática vêm se delineando ao longo do tempo, pretende-se, também, identificar quais são os autores mais influentes e os mais produtivos. O objetivo geral é apresentar o panorama da produção científica sobre crítica diplomática a partir da análise de dados coletados nas bases de dados Google Scholar, Web of Science e Scopus.

2. Caminhos metodológicos

Baseado no método utilizado por Diaz, Alvarez e Vidotti (2018), o estudo, de natureza exploratório-descritiva, foi realizado a partir de uma análise bibliométrica, objetivando obter uma dimensão da

produção científica no tema da Crítica Diplomática. Segundo Oliveira (2018), os estudos métricos da informação constituem um dos instrumentos metodológicos consistentes para a visualização do comportamento da ciência em um dado campo.

A coleta de dados foi realizada janeiro de 2022, utilizando como fonte o software Publish or Perish (PoP), e as bases selecionadas para a coleta de dados foram Google Scholar, Web of Science e Scopus. Como estratégia de busca, foram definidas a expressão “*diplomatic criticism*” e suas traduções “crítica diplomática” e “*critique diplomatique*” em espanhol e francês, respectivamente; foram realizadas buscas simples nos campos *keyword* e *title words*, sem limitação de data de publicação. Utilizou-se o software Publish or Perish (PoP) por motivo de sua interface emitir relatórios que facilitam a análise bibliométrica pretendida.

Para este estudo, foram utilizados os indicadores Bibliométricos a seguir, conforme o Quadro 1.

Indicadores bibliométricos	Descrição
Indicadores de produção	São constituídos pela contagem do número de publicações e objetivam refletir seu impacto junto à comunidade científica à qual pertencem, dando visibilidade àqueles mais produtivos e às temáticas mais destacadas de uma área de conhecimento.
Anos de publicação	Número da produção por anos e séculos
Quantidade de trabalhos produzidos	Número total de resultados
Anos versus trabalhos produzidos	Gráfico comparado dos anos x produção
Indicadores de Impacto e visibilidade	Permite a identificação de grupos de cientistas e suas publicações, com finalidade de evidenciar os pesquisadores de maior impacto
Quantidade de citações recebidas	A soma das contagens de citações em todos os resultados atualmente selecionados
Média de citações por ano	Número médio de citações por ano (ou seja, citações/anos de citação)
Média de citações por documento	A soma das contagens de citações em todos os artigos dividida pelo número total de artigos. A média e o modo também são calculados e disponibilizados separadamente
Citações por autor	Número médio de citações por autor
Artigos por autor	Número médio de artigos por autor
Autores por documento	Número médio de autores por artigo
H-index	Índice H de Hirsch
Índice g	Índice g de Egghe

Quadro 1. Descrição dos indicadores bibliométricos analisados (adaptado de Oliveira, 2018; Diaz, Alvarez e Vidotti, 2018)

De acordo com o método utilizado por Diaz, Alvarez e Vidotti (2018, p. 149), o uso da ferramenta PoP para coleta e processamento, além de auxiliar a construção dos gráficos para análise dos dados, emite relatórios em .csv e possibilita recuperar citações, citações por autor, ano, tipo de documento e resumos. A ferramenta traz também o ranking de citações do Google com base nas citações registradas pelos robôs de busca dos indexadores Google, além de recuperar documentos nas seguintes línguas: árabe, polonês, japonês, sérvio, alemão, russo, tcheco, coreano, chinês e ucraniano, mesmo não sendo pesquisado em tais línguas.

Após a coleta de dados, o uso do programa PoP requer a limpeza da lista de publicações obtidas na busca, quando também são recuperados registros que não resultaram em produção científica, como planos de ensino e projetos de curso vinculados à disciplina diplomática em algumas universidades internacionais. Além disso, o PoP limita seus resultados a mil registros, o que não comprometeu a coleta, pois a pesquisa resultante em mil registros foi delimitada por tempo em séculos e, em alguns casos, por décadas (após 2000). Dessa forma, a pesquisa recuperou 2.758 documentos e após a limpeza de dados retirando os duplicados obtivemos 1.648 documentos.

Para um cálculo mais aproximado do real, foi realizada a limpeza de dados, retirando-se da planilha Excel as informações com duplicidade ou aquelas que não iriam contribuir para o estudo devido à recuperação de documentos tais como planos de ensino e documentos inerentes a Projetos dos cursos de Arquivologia por não se enquadrarem em produção científica.

Os resultados obtidos no PoP foram salvos em formatos .csv e compilados em planilha Excel, e na limpeza dos dados foram utilizados recursos para arrumação dos dados visando à padronização da entrada de autores, no qual se utilizou a função

```
=CONCATENAR(DIREITA(A1;NÚM.CARACT(A1)-
LOCALIZAR(" ";A1;1));";ESQUERDA(A1;
LOCALIZAR(" ";A1;1)-1))
```

para reorganizar os nomes dos autores no padrão (SOBRENOME, NOME), a fim de facilitar na confecção de tabelas dinâmicas dentro do software Excel.

3. Resultados

De acordo com o levantamento realizado nesta pesquisa, os estudos em Crítica Diplomática possuem cerca de três séculos e meio de pesquisas e publicações. A primeira publicação disponível data do início, em 1681, com a publicação considerada a matriz do método, o livro “De re diplomatica libri VI”, de Jean de Mabillon que consiste em uma coletânea de 6 volumes.

Foram recuperados um total de 2.758 documentos, sendo 12 na WoS, 22 na Scopus e 2.724 no Google Scholar, por meio do software Hazing – PoP, distribuídos conforme o Gráfico 1 ao longo dos séculos, com um aumento significativo na produção no fim dos anos de 1990, permanecendo em crescimento até o momento.

De 1681 a 1968, a produção científica sobre Crítica Diplomática se manteve estável na quantidade de 1 a 5 publicações por ano, massivamente na Europa, e nos séculos XX e XXI, é possível observar um aumento considerável na produção científica sobre Crítica Diplomática. Esse aumento pode ser justificado pela necessidade de estudos em torno dos documentos eletrônicos e digitais, assim como pela iniciativa do projeto Interpares (3), coordenados pela pesquisadora Luciana Duranti, que buscou desenvolver o conhecimento teórico-metodológico essencial para a preservação de documentos arquivísticos digitais autênticos. A nuvem de palavras foi construída utilizando-se a análise das palavras-chave dos artigos em formato .txt e a visualização no site makewordcloud.com.

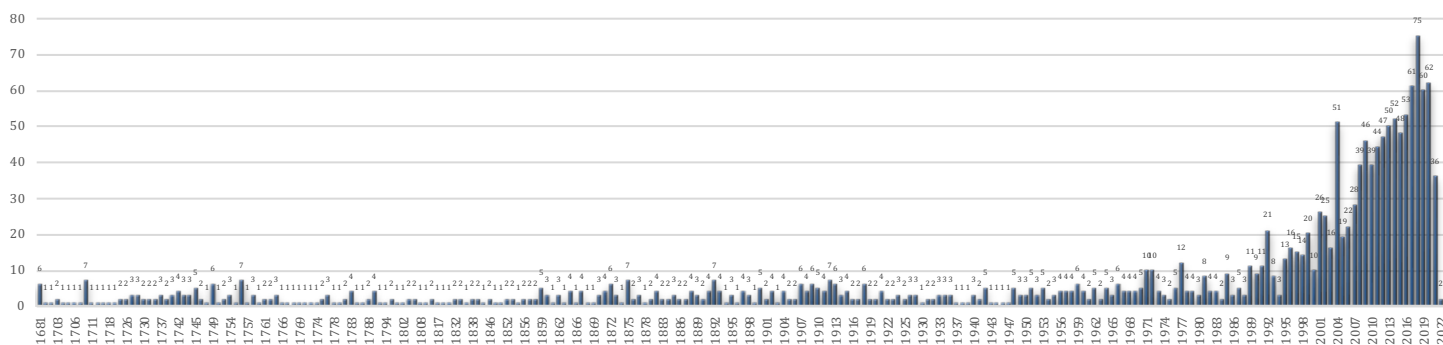


Gráfico 1. Gráfico da visualização da produção científica na temática Crítica Diplomática (adaptado de Oliveira, 2018; Diaz, Alvarez e Vidotti, 2018)

3.1. Produção dos autores

Com a produção na temática “Diplomática” e “Crítica Diplomática” recuperada pelo PoP, mensurando os 10 autores mais produtores, tem-se quadro 2, na qual constam autores e pesquisadores de diferentes países e instituições na mesma linha de autores contemporâneos e autores clássicos da literatura sobre Crítica Diplomática.

O levantamento resultou em 1.507 autores que produziram 1.648 pesquisas. Dos 1.507 pesquisadores, 1.295 publicaram somente um trabalho, o que corresponde a 85,93%, e 212 publicaram acima de 2 trabalhos, correspondendo a 14,07%.

Autores	N
Tognoli, Natália Bolfarini (Universidade Federal Fluminense-UFF/ Brasil)	16
Mabillon, Jean (Collège Université de Reims - França)	12
Guimarães, José Augusto Chaves (Universidade Estadual Paulista - UNESP - Brasil)	10
Duranti, Luciana (University of British Columbia - Canadá)	9
Escalona, Julio (University College London – Reino Unido)	9
Von Teutschbrunn, Johann Heumann (Universidade de Altdorf - Alemanha)	8
Rodríguez, Sonia Maria Troitíño (Universidade Estadual Paulista-UNESP/Brasil)	7
Barone, Nicola (Grande Arquivo de Napoli – Itália)	6
Shchekochikhina, MA (Universidade Estadual de Moscou – Rússia)	6
Rabello, Rodrigo (Universidade de Brasília – Brasil)	5
Andrade, Elias Alves (Universidade Federal do Mato Grosso – Brasil)	4

Quadro 2. Autores mais produtivos em Diplomática

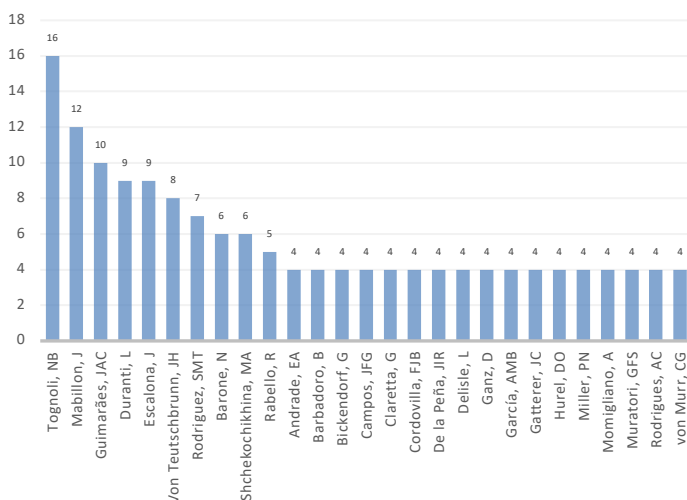


Gráfico 2. Autores mais produtivos a nível mundial

Na análise do gráfico 2, destaca-se que 212 (14,07%) dos autores mais produtivos foram responsáveis por 567 (34,41%), das pesquisas do total de 1.648 publicadas. Esse percentual sugere a significativa representatividade do grupo mais produtivo.

Destaca-se que o levantamento aqui realizado só foi possível porque os materiais foram escaneados pelo Projeto Google Books para serem tornados público no Google Library Search. Porém, devido a problemas com direitos autorais, o projeto foi descontinuado em alguns países como a França e Reino Unido.

Assim, a digitalização permite o acesso independentemente da distância geográfica, possibilitando, dessa forma, acesso mundial através da Web.

Indicadores de produção	
Anos de publicação	1681-2021
Quantidade de trabalhos produzidos (até janeiro de 2022)	1648
Anos de citação (até janeiro de 2022)	341 anos (1681-2022)
Indicadores de impacto e visibilidade	
Quantidade de citações recebidas	15723
Média de citações por ano	46,10
Média de citações por documento	9,54
Citações por autor (4)	10,43
Documentos por autor (5)	1,09
Autores por documento	1,13
H-index	41
Índice g	107

Quadro 3. Indicadores bibliométricos da temática diplomática

A quantidade de anos de publicação de uma temática demonstra a solidez e relevância dos estudos e a maturidade das teorias em torno dela, visando conhecer os principais teóricos e compreender o conjunto de autores que contribuem significativamente para a disciplina em estudo.

Quanto aos indicadores de impacto e visibilidade, as citações são entendidas como parte dos sistemas de recompensa e reconhecimento existentes (como os prêmios e bolsas), que atuam ajustando o comportamento dos investigadores. O crédito dado a um pesquisador por outro, por meio de uma citação, representa uma forma de reconhecimento (Romancini, 2010, p. 22).

Concordando com Leydesdorff (1998), entendemos o estudo de citação como algo que precisa ser explicado *explanandum*, e não apenas como

um indicador que busca expressar algo (o crescimento da ciência, o surgimento de especialidades, a consolidação e expansão de disciplinas

etc.), ou seja, um *explanans*. É isso se busca na análise dos autores mais citados.

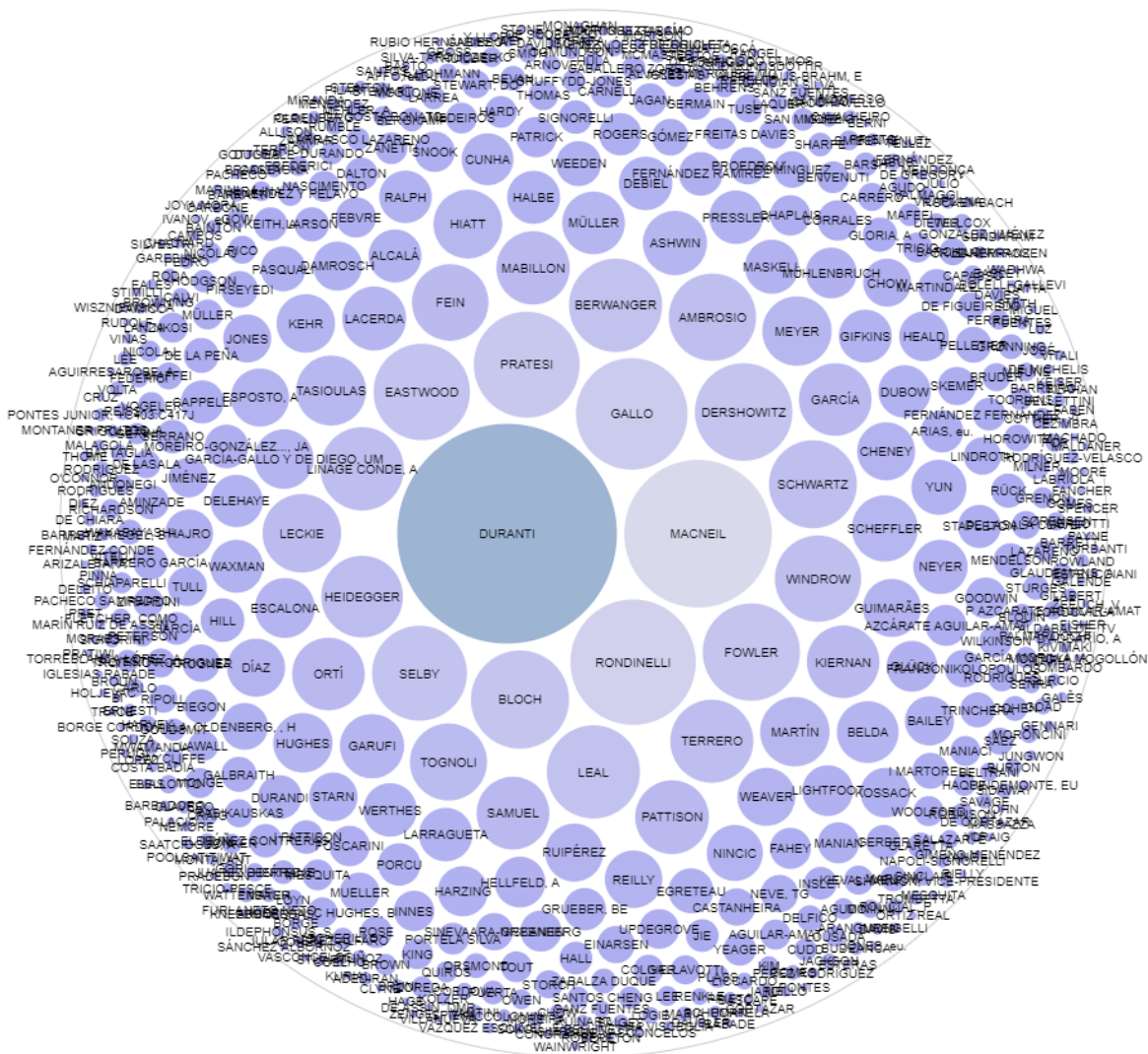


Figura 1. Visualização dos autores e a métrica do total de citações recebidas

A Figura 1 traz os autores mais citados no centro da imagem, o que significa que quanto maior o número de citações, maior o raio de cobertura dos autores que estão localizados na imagem. Os autores mais citados nos estudos sobre Crítica Diplomática e ou Diplomática são:

Ordem	Autores	Citação
1°	DURANTI, L	862
2°	MACNEIL, H	389
3°	WATTENBACH, W	332
4°	RONDINELLI, RC	281
5°	GALLO, AG	222

6°	PRATESI, A	200
7°	BLOCH, M	172
8°	DERSHOWITZ, AM	170
9°	FOWLER, DC	169
10°	EASTWOOD, T	166
11°	SELBY, J	161
12°	BERWANGER, AR	155
13°	LEAL, JEF	155
14°	SCHWARTZ, JM	138
15°	AMBROSIO, T	136
16°	WINDROW, M	133
17°	LINAGE CONDE, A	124

18°	TERRERO, AR	123
19°	HEIDEGGER, M	119
20°	LECKIE, S	110
21°	TOGNOLI, NB	106
22°	FEIN, H	105

Quadro 4. Autores mais citados na temática

A pesquisadora Luciana Duranti possui nove trabalhos que são resultados de pesquisas iniciadas há 32 anos, e a maioria está disponível em acesso aberto. Entre os trabalhos recuperados, o mais citado é *Diplomatics: new uses for an old science* (Duranti, 1989), primeiro de uma série de seis artigos publicados no periódico *Archivaria*, da Associação dos Arquivistas Canadenses. Antecedido por uma apresentação que o contextualiza, a autora aborda, neste artigo, o que é a Diplomática, a história dessa ciência, seu objeto e objetivos e discute os conceitos de autenticidade e originalidade, enfatizando a importância do uso dos métodos diplomáticos para o trabalho arquivístico nas atividades de identificação, avaliação, arranjo e descrição de documentos. Neste artigo, publicado em 1989, mas só traduzido para o Brasil em 2020, a autora conclui que as características dos documentos arquivísticos contemporâneos requerem uma análise baseada numa integração da teoria da Diplomática com a teoria da Arquivística, porque, segundo ela, nenhum documento arquivístico hoje se mantém isolado à maneira como os medievais o fizeram, e o seu significado é determinado pelo seu contexto documental, para além dos seus contextos administrativos e jurídicos.

Macneil, H. possui, nesse levantamento, três trabalhos, todos e-books disponíveis para aquisição pela Springer. Um deles, o livro *Registros de Confiança: perspectivas jurídicas, históricas e diplomáticas* explora a evolução dos princípios e métodos para determinar a confiabilidade do documento da antiguidade na era digital e das perspectivas do direito e da história, sendo uma obra de interesse de estudiosos jurídicos que atuam no campo do direito das evidências, historiadores que atuam no campo da metodologia histórica e profissionais de registro (gerentes de registros, especialistas em tecnologia da informação, arquivistas) que trabalham na concepção e implementação de sistemas de registro organizacional contemporâneos. A segunda obra mais citada é o livro em coautoria com Luciana Duranti e Terry Eastwood intitulado *Preservation of the integrity of electronic records*, citada mais de 166 vezes.

A obra de Wattenbach (1871), escrita em alemão e intitulada *Das Schriftwesen im Mittelalter*, cuja

tradução é *Escrevendo na Idade Média*, é um livro digitalizado pelo Google e disponível gratuitamente para download; obra rara, mas em bom estado, apesar de escrita há três séculos. Considerado um clássico, o livro relata os primórdios da Diplomática e é voltado para paleógrafos, mas atende a todos os interessados na história das escrituras de Idade Média. No terceiro capítulo, o autor trata das falsificações de documentos medievais e de diplomática sem o uso do termo diplomática.

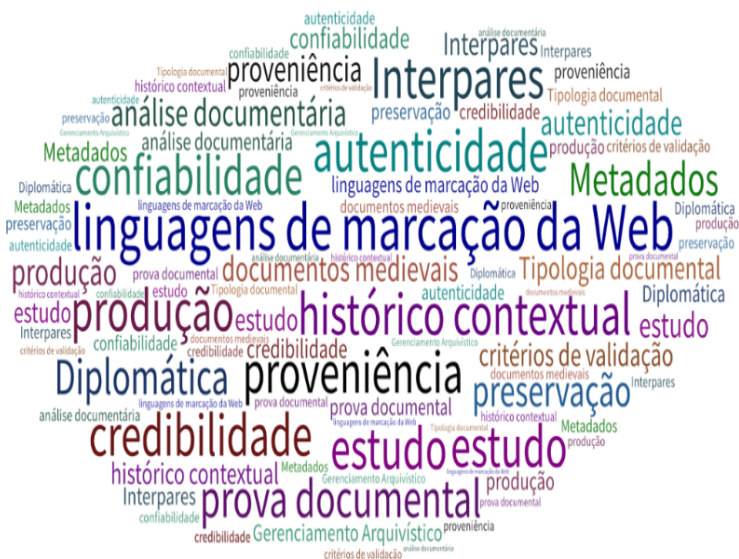


Figura 2. Visualização da nuvem de palavras

O principal foco das pesquisas desenvolvidas no cenário mundial aborda os estudos sobre a diplomática enquanto metodologia, visando garantir autenticidade, fidedignidade e preservação de documentos arquivísticos para identificação em Sistema Eletrônico de Informações, além de sua relação com a custódia de documentos, entre outros que estão representados na nuvem de palavras da Figura 2.

4. Considerações finais

Conclui-se que, mesmo sendo uma ciência que se originou em meados de 1681, o uso de softwares para extração e visualização de dados nas análises bibliométricas é comum na área de Ciência da Informação, embora as configurações dos modelos complexos e altamente personalizados ainda sejam difíceis para os profissionais que não possuem as competências necessárias durante o desenvolvimento da pesquisa, o que demandaria uma quantidade significativa de tempo e estudo para correto domínio das ferramentas.

Salienta-se que as principais publicações recuperadas sobre o tema estudado não são oriundas de fontes ou tipologias documentais tradicionais de análise métrica da produção científica, e que os livros são as principais fontes do foco de estudo, de acordo com as métricas de citação.

Comportamentos interessantes foram identificados em relação às métricas de citações recebidas de artigos científicos de fontes alternativas à de sua publicação original, como é o caso das redes sociais acadêmicas (academia.edu, ResearchGate) que apresenta indicadores dos coautores das publicações. Este fato demonstra a tendência dos pesquisadores em compartilhar suas publicações e ganhar maior visibilidade para seu trabalho nessas plataformas, para além da publicação em periódicos científicos especializados.

O ano de 2018 foi identificado como o ano mais produtivo em produções científicas Crítica Diplomática, com 56 publicações, seguido pelo ano de 2016, com 54 publicações. Em resumo, mais atual do que nunca, a avaliação e mensuração da produção científica sobre a Diplomática/Crítica Diplomática constitui-se atividade necessária e constante. Os resultados aqui apresentados mostram uma tendência promissora de desenvolvimento desse campo de estudo.

Notas

- (1) O Projeto Google Biblioteca, iniciou em 2004, e em 2007 foi ampliado para 28 parceiros, incluindo sete bibliotecas internacionais, das quais: Universidade de Oxford (Reino Unido), Universidade Complutense de Madri (Espanha), a Biblioteca Nacional da Catalunha (Espanha), a Biblioteca da Universidade de Lausanne (Suíça), a Universidade de Ghent (Bélgica) e a Universidade de Keio (Japão), além das Bibliotecas de Harvard, Michigan, Biblioteca Pública de Nova York, Biblioteca da Universidade de Stanford entre outras, para mais informações a respeito consultar: Projeto Google Books Library.
- (2) O "CSV" em um arquivo de aposta com o .csv extensão significa "valores separados por vírgula", porque os dados desses arquivos CSV são detalhes divididos por vírgulas em conjuntos específicos de informações. Esses pedaços de dados podem ser inseridos por usuários de aplicativos de planilha e de edição de texto integrado com suporte para a criação e modificação de documentos CSV.
- (3) O projeto Inter pares visa desenvolver o conhecimento essencial para a preservação a longo prazo de registros autênticos criados e/ou mantidos em formato digital e fornecer a base para padrões, políticas, estratégias e planos de ação capazes de garantir a longevidade de tal material e a capacidade de seus usuários confiarem em sua autenticidade. Os resultados e produtos das três primeiras fases do projeto podem ser encontrados no site: <http://interpares.org/>.
- (4) Para cada trabalho, a contagem de citações é dividida pelo número de autores para esse artigo, para devolver a contagem de citações normalizada por autor para o artigo. As contagens de citações normalizadas são então somadas e, todos os artigos para fornecer o número de citação por autor sobre o conjunto de dados.
- (5) Para cada trabalho, a contagem de 1 / autor é calculada para fornecer a contagem de autor normalizada para o trabalho. A contagem normalizada dos autores é então resumida em todos os artigos para fornecer o número de artigos por autor.

Agradecimientos

Agradecimentos à Fundação de Amparo à Pesquisa do Estado do Piauí (FAPEPI) pelo apoio financeiro referente ao Edital nº 001/2022 Processo nº 00110.000156/2022-67.

Referências

- Bellotto, H. P. (2002). Como fazer análise diplomática e análise tipológica de documento de arquivo. São Paulo: Arquivo do Estado de SP, Imprensa Oficial. http://www.arquivoestado.sp.gov.br/site/assets/publicacao/anexo/co_mo_fazer_analise_diplomatica_e_analise_tipologica.pdf (2022-12-22).
- Díaz, M. P.; Alvarez, E. B.; Vidotti, S. A. B. G. (2018). Humanidades digitais: visualização da produção científica. // Workshop de Informação, Dados e Tecnologia, Paraíba. <https://www.ufpb.br/widat/contents/documentos/anaiswidat2018.pdf#page=147>(2022-12-22).
- Duranti, L. (1995). Ciência archivística. Trad. Manuel Vazquez. Córdoba, Argentina, 1995.
- Duranti, L. (2015). Diplomática: novos usos para uma antiga ciência (parte v). // Acervo - Revista do Arquivo Nacional. 28:1, 196-215, 1995
- Harzing, A. et al. (2011). The Publish or perish book. Melbourne: Tarma Software Research, 2011.
- Leydesdorff, L. (1998). Theories of Citation? // Scientometrics. 43:1, 5-25, 1998.
- Oliveira, E. F. T. (2018). Estudos métricos da informação no Brasil: indicadores de produção, colaboração, impacto e visibilidade. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2018.
- Romancini, R. (2010). O que é uma citação? A análise de citações na ciência. // Intexto. Porto Alegre. 23:2 (2010) 20-35, 2010.
- Tallafigo, M. R. (1994). Ayer y hoy de la diplomática, ciencia de la autenticidade de los documentos. // Carucci, P. et al. Documento y archivo de gestión: diplomática de ahora mismo. Carmona: Universidade Internacional Menéndez Pelayo, 1994.
- Tessier, G. (1952). La diplomatie. Paris: Press Universitaires de France, 1962.
- Valente, J. A. V. (1978) Acerca do documento. // Revista Brasileira de Biblioteconomia e Documentação. 11:3/4, 177-198, 1978.
- Wattenbach, W. (1871). Das Schriftwesen im Mittelalter. Hirtzel. <https://books.googleusercontent.com/books/content?req=AKW5QaeKalu> (2022-12-22).

Enviado: 2023-03-29. Segunda versión: 2023-06-09.
Aceptado: 2021-06-16.

El artefacto como sistema generador de comunidades epistémicas en ambientes científicos de la nanotecnología

The artefact as a generating system of epistemic communities in scientific environments of nano-technology

Heriberto RAMÍREZ, Javier TARANGO, Juan D. MACHIN-MASTROMATTEO

Universidad Autónoma de Chihuahua, Avenida Universidad s/n, Chihuahua, México 31174.
hramirez@uach.mx, jtarango@uach.mx, jmachin@uach.mx

Resumen

La mayor parte del equipamiento del mundo está compuesto por artefactos: edificios, carreteras, redes eléctricas, antenas de telecomunicación, automóviles, viviendas, teléfonos móviles, por mencionar solo algunos. En su conjunto, desde la antigüedad se han ido creando, principalmente con fines de mejorar las condiciones de vida de las personas. Sin embargo, es en épocas recientes que este mobiliario se ha convertido en objeto de estudio para su análisis filosófico, histórico y sociológico. Este artículo se divide en tres partes: (1) un recuento histórico de los aparatos epistémicos, especialmente relacionado con los microscopios; (2) una revisión que ofrece una panorámica general de la filosofía del artefacto y su relación con los estudios de la información; y (3) un estudio de caso, centrado en el intento por desvelar la cultura epistémica de la comunidad de nanotecnólogos de un centro público de investigación (Centro de Investigación en Materiales Avanzados, CIMAV, Chihuahua, México) y cómo se valen de un microscopio electrónico en su modalidad de barrido y transmisión, en la generación de conocimiento a través de acciones sustanciales en el desarrollo de publicaciones científicas, formación de recursos humanos, participación sistemática de investigadores en proyectos rentables y formación de comunidades epistémicas. Este estudio identifica la ausencia de modelos específicos para una precisa medición de la efectividad de los artefactos y la necesidad de desarrollar propuestas formales que fomenten la relación en las actividades científicas basadas en artefactos.

Palabras clave: Filosofía de los artefactos. Filosofía de la tecnología. Filosofía de la técnica. Práctica científica. Maquinaria epistémica. Artefacto epistémico. Generación de conocimiento. Comunidades epistémicas. Centro de Investigación en Materiales Avanzados (Chihuahua, México). México.

1. Introducción

A lo largo de la historia hemos amueblado el mundo con un sinfín de artefactos. Aunque pocas veces nos preguntamos acerca de su naturaleza y menos acerca de sus posibles efectos sobre la humanidad y su visión del mundo. La filosofía de la tecnología y del artefacto se hace estas y otras

Abstract

Most of the world's equipment is made up of artifacts: buildings, roads, power grids, telecommunication antennas, automobiles, homes, mobile phones, to name just a few. As a whole, since ancient times they have been created, mainly for the purpose of improving people's living conditions. However, it has not been until recent times that this furniture has become an object of study for its philosophical, historical and sociological. This article is divided in three parts: (1) a historical account of epistemic devices, especially related to microscopes; (2) a revision aimed at offering a concise overview of the philosophy of the artifact and its relationship with information studies; and (3) a case study, focused on the attempt to reveal the epistemic culture of the community of nanotechnologists from a public research center (Centro de Investigación en Materiales Avanzados, CIMAV, Chihuahua, Mexico) and how they use a microscope. electronic in its sweep and transmission modality, in the generation of knowledge through substantial actions in the development of scientific publications, training of human resources, systematic participation of researchers in cost-effective projects and for developing epistemic communities. This study identifies the absence of specific models for a precise measurement of the effectiveness of artifacts and the need for developing formal proposals to promote the relation with scientific activities based on artifacts.

Keywords: Philosophy of artifacts. Philosophy of technology. Philosophy of technique. Scientific practice. Epistemic machinery. Epistemic artifact. Knowledge generation. Epistemic communities. Centro de Investigación en Materiales Avanzados (Chihuahua, Mexico). Mexico.

preguntas. El andamiaje que ha construido esta joven rama de la filosofía, se ha encaminado a dilucidar sobre los problemas filosóficos de la tecnología, lo cual invita, en primera instancia, a tender a ser perspicaces en esta vertiente distintiva entre la filosofía de ingenieros y la filosofía de humanistas, puntualizada en la difundida obra de Mitcham (1989) y ampliada más adelante en otra

de sus obras (Mitcham, 1994), recuperada reciente y oportunamente por Quintanilla (2020).

A partir de esto, conviene caracterizar la diferenciación entre las filosofías mencionadas, pues cada una tiene que ver con otras dos nociones importantes: (1) cuando la visión ingenieril se asocia con la idea del cambio tecnológico, su semántica suele incluir el diseño y la construcción de artefactos; y (2) cuando se ve en conjunción con la evolución tecnológica y se establece una relación con “la dinámica interna de la trama de los objetos técnicos que componen el sistema técnico de una época” (Sandrone, 2022, p. 364). En nuestro mundo superpoblado de artefactos surgen las preguntas: ¿Cómo estudiarlos? ¿Qué conceptualización se ha sugerido para su abordaje? En un primer momento podría considerarse su composición material, de qué están hechos, cómo fueron producidos; un segundo momento, tiene que ver para qué fueron hechos, cuál es su función o utilidad o su historia intencional. Suele aceptarse esta doble vía de abordaje, la cual está ya considerando una naturaleza dual de los artefactos.

Si bien el mundo artefactual está sobrepoblado, a tal extremo que se nos ha vuelto cada día más difícil distinguir el mundo natural del artificial, vislumbrándose cada vez con mayor cercanía las ciencias de lo artificial. Aquí, el estudio se centra en una clase específica de artefactos: los epistémicos, los cuales se definen como aquellos diseñados para generar conocimientos e información confiables, por lo tanto, son artefactos que contribuyen a establecer los estándares del conocimiento. Para el caso de este estudio, en específico, se consideran como objeto de estudio aquellos concebidos para explorar el mundo de lo más pequeño o la ciencia nanotecnológica.

La curiosidad puede ser la clave de acceso para desvelar los misterios del mundo natural o social, alimentada por la generación de preguntas creativas y desafiantes: ¿el recorrido histórico de los artefactos vinculados con los microscopios ofrece antecedentes suficientes para definir su condición actual en el ambiente científico y nanotecnológico? ¿Cuál es la forma en que los artefactos funcionan como medios de consulta y generación de conocimiento a través de la integración de comunidades epistémicas? ¿Las condiciones de trabajo del centro de investigación en estudio, permiten definir sus condiciones de rentabilidad en la generación de diversas actividades vinculadas con la producción científica? Para ello, este texto planea como objetivo realizar los siguientes abordajes:

1. Integrar un recorrido histórico somero, el cual incluye la mención de algunos artefactos sig-

nificativos, asociados en su mayoría a la naturaleza de la luz y conducentes, así como la composición de la materia.

2. Analizar la filosofía de los artefactos, de la técnica y de la tecnología como agentes de cambio de paradigma en las formas de consulta y generación de conocimiento a través de diversos medios de información decodificada y la generación de otros más usando información codificada, así como su vinculación con los estudios de la información.
3. Examinar un caso sobre artefactos epistémicos (microscopio de barrido y microscopio de transmisión) en cuanto a condiciones de medición de su efectividad en el Centro de Investigación en Materiales Avanzados (CIMAV), centro público de investigación especializado en físico química de materiales, ubicado en la ciudad de Chihuahua, México.

2. De la filosofía de la ciencia a la filosofía de los artefactos

Los efectos del avance científico-tecnológico actual resultan un tema prioritario en el estudio de los artefactos, cuya naturaleza representa un punto de análisis interesante, el cual es posible desde múltiples enfoques y perspectivas (Márquez, 2020) que se resumen en la Figura 1, en la página siguiente. El estudio de la filosofía de los artefactos, como una parte fundamental de la filosofía de la ciencia, demanda una clasificación exhaustiva de múltiples elementos que los distinguen según su estructura, función, intención y uso, además, se vuelven necesarios como referencia, los propósitos para los que fueron creados y, por tanto, sus aportaciones a la solución de problemas o contribuciones a la generación de una cultura de la información en los actores que se involucran con ellos, a través de dos condiciones: (1) los elementos materiales que los distinguen; y (2) sus condiciones normativas a través de sus propiedades intrínsecas, de respuesta y técnicas (Broncano, 2008). En la filosofía de los artefactos no se pueden obviar las valoraciones antropocéntricas basadas en la utilidad, función, poder simbólico o eficiencia que el propio hacedor o realizador otorga a los artefactos creados (Parente, 2020).

Se defiende que los artefactos realizan funciones para satisfacer objetivos relacionados con su diseño, producción y comercialización (Lawler, 2003). En esta visión antropocéntrica, el usuario o investigador toma a los artefactos como: (1) medios sin genuina acción y neutrales en cuanto a valores, sometidos a ellos mismos para el cumplimiento de sus propósitos; (2) depositarios genui-

nos de valores sustantivos porque objetivan intenciones y formas de vida, intrínsecamente portan valores culturales; y (3) reconceptualizan la acción

en cuestiones de mediación técnica para la consecución de un fin (Vega Encabo, 2009), por tanto, la función del artefacto puede ser analizada desde un sentido histórico y no histórico.

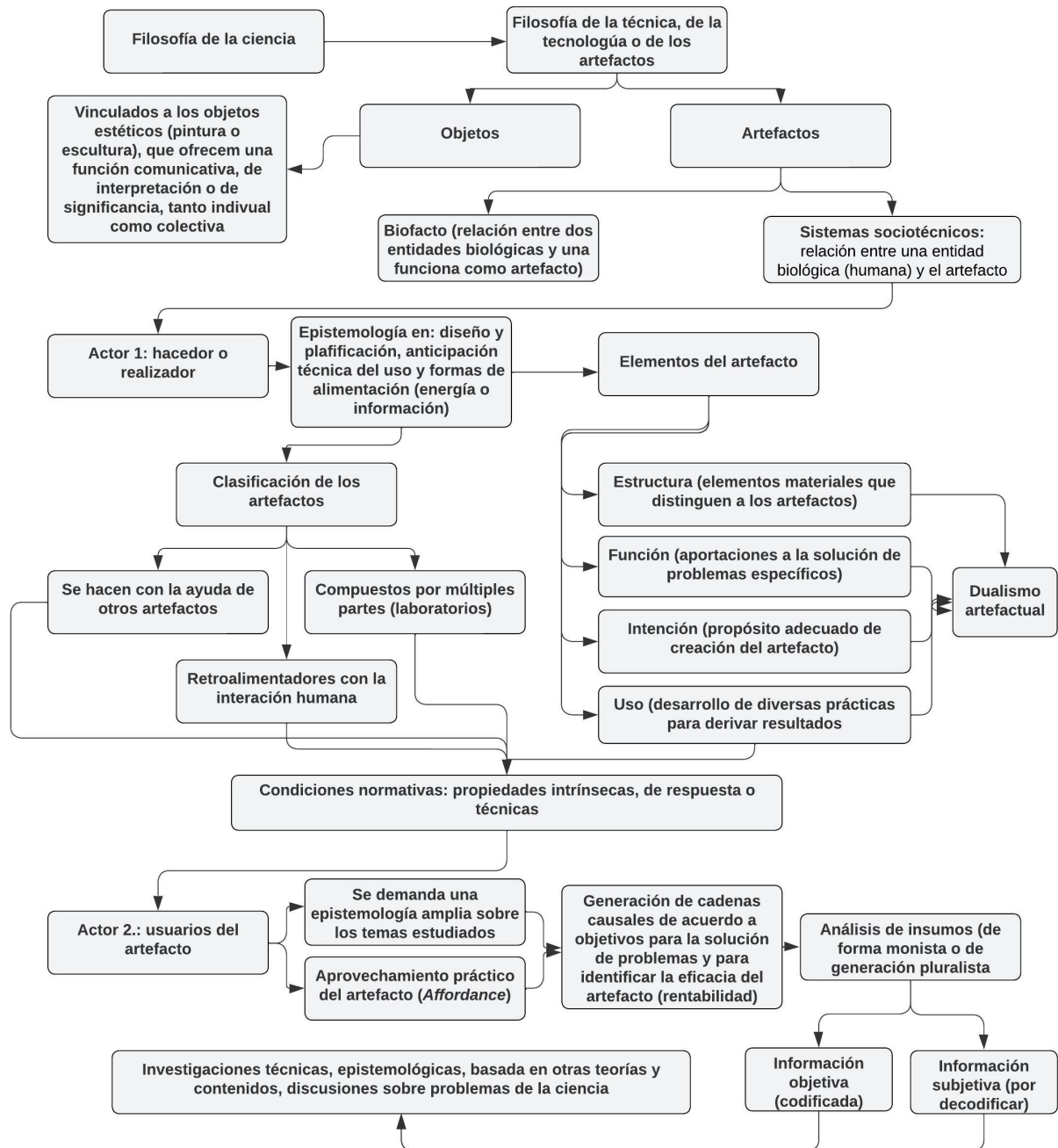


Figura 1. Procesos y actores en la filosofía de los artefactos

La definición de los artefactos como objetos de estudio de la filosofía de la ciencia y tecnología se refiere a una variedad inmensa de cosas, tales como: herramientas, documentos, joyería, instrumentos científicos, máquinas, muebles, u objetos (Asse, 2015). No obstante, resulta poco probable

la generación de una teoría unificada que explique las características de cada uno de los ejemplos enlistados, resultando en la constitución de sistemas híbridos o sociotécnicos, según el propósito tanto del artefacto como del individuo de cada uno (Mejía Rendón, 2023). Esta propuesta

sobre la hibridación estará siempre constituida por humanos y ambientes artificiales (no humanos), imbrica habilidades humanas y un andamiaje artificial, es el justo medio de las intenciones, prácticas y capacidades cognitivo-agenciales (Parente, 2020). Debe considerarse en este aspecto que la hibridación también puede suceder de forma no convencional entre dos entidades biológicas, conocido esto como *Biofacto* (contracción entre una entidad biológica y un artefacto).

En la intención por definir de manera más precisa los aspectos que caracterizan a los artefactos producto de la técnica, Lee (2020) y Mejía Rendón (2023) los resumen en los siguientes puntos: (1) observan características físicas (masa, forma, altura, velocidad o energía); (2) son producidos por un hacedor bajo determinadas intenciones; (3) son concebidos para realizar determinadas funciones; (4) presentan elementos normativos para un uso correcto y no necesariamente incorrecto; y (5) cumplen con funciones sociales y son reconocidos bajo diversos elementos simbólicos. Monterroza-Ríos et al. (2015) complementan esta propuesta a través de tres argumentos de la filosofía de la técnica, la tecnología y el artefacto: (1) orden jerárquico; (2) neutralidad axiológica; e (3) instrumentos de poder y dominación.

La clasificación de los artefactos suele ser variada y compleja, según los propósitos por los que fueron creados, ofreciendo características individuales y combinadas a través de propuestas duales. Una aproximación a tal taxonomía considera que los artefactos se deben clasificar según su naturaleza: (1) por las estructuras físicas en que fueron diseñados; (2) por el cumplimiento de ciertas funciones que refieren a la intencionalidad humana (Broncano, 2008; Monterroza-Ríos, 2010); y (3) no se establece la distinción entre función y estructura, la cual es central en el dualismo artefactual (Parente, 2008; Peterson y Spahn, 2011; Vaccari, 2011) y es una característica llamativa. Se resume esta clasificación en función de tres conceptos: (1) intencionalidad del artefacto o el objetivo por el cual fue creado; (2) funcionalidad del artefacto o el uso para desarrollar ciertas funciones en un determinado grupo cultural; y (3) el enfoque dual que reconoce que los artefactos son entidades materiales e intencionales (Monterroza-Ríos, 2011).

La naturaleza dual de los artefactos ha sido analizada por el Programa de Naturaleza Dual de los Artefactos Técnicos (Dual Nature of Technical Artifacts Program), proyecto basado en el intencionalismo que surge en una universidad tecnológica de Holanda y defiende como tesis central que los artefactos se constituyen de dos elementos: físicos e intencionales, cuyas funciones ofrecen condiciones ontológicas que permiten distinguir

asignaciones correctas e incorrectas, todo ello a través de la aplicación de disciplinas filosóficas y de las ciencias sociales en relación con la ingeniería, el diseño y la tecnología (Monterroza-Ríos, 2010; Vaccari, 2011)

En relación con las estructuras físicas de los artefactos y la forma como físicamente están constituidos, estas pueden a su vez dividirse bajo las siguientes características: (1) artefactos que se hacen con la ayuda de otros artefactos; (2) artefactos compuestos por múltiples partes; y (3) la interacción con los artefactos incide en el desarrollo de técnicas humanas. Así también, en relación con la intencionalidad humana, esta a su vez, puede estar compuesta por dos elementos: (1) el sustento cognitivo a través de esquemas mentales del hacedor para crear y mejorar los artefactos usando la creatividad; y (2) las características motivacionales del hacedor de artefactos, que no precisamente tienen relación con necesidades físicas o biológicas de supervivencia (Monterroza-Ríos, 2019).

A través de los artefactos se logra una fecunda transmisión cultural, sin embargo, tal cultura es solo un aspecto del proceso, incluso se consideran de mayor relevancia las aportaciones que los artefactos ofrecen para transformar el medio, pasando de las posibilidades imaginarias del sujeto (hacedor) que los creó y que luego se convierten en posibilidades reales (Broncano, 2008). Todo ello tiene sus orígenes en la posición aristotélica de los artefactos, basada en tres principios: (1) carecen de esencia; (2) carecen de un principio común de actividad, normas para comenzar o dejar de existir; y (3) se agrupan según su función y no según su constitución. Los artefactos se distinguen por sus "accesos prácticos" conocidos como andamiajes (*affordances*) o aprovechamiento práctico percibido (Vega Encabo, 2009). El instrumentalismo distingue entre "sujeto" (usuario activo independiente) y "objeto" (artefacto pasivo dependiente) (Parente, 2020). Otra distinción importante dentro de la filosofía de los artefactos, de la tecnología y la técnica es la diferenciación entre órgano y útil (basada en la visión del filósofo Heidegger), esta diferenciación de interrelación entre lo orgánico (ser humano) y lo artificial (artefacto), propician su uso de acuerdo con el estatuto ontológico con que fue creado el artefacto (Parente, 2008).

La filosofía de los artefactos no cobraría relevancia sin la teoría de los hacedores, provocando con ello una alternativa al modelo dominante sobre la ontología de los artefactos. Para ello, se considera que se entienden más los artefactos a través de sus hacedores, en lugar de solo conocer su funcionalidad, provocándose con ello la búsqueda

en la diferenciación entre los conceptos de realizador y realización, así como la realizabilidad múltiple (Parente y Vaccari, 2022). En relación con el conocimiento, la interacción realizador-realización implica: (1) el conocimiento expresado en habilidades reclama una epistemología personal (el conocimiento explícito tiene algo de conocimiento tácito); (2) el conocimiento práctico puede expresarse como pensamiento visual; y (3) el conocimiento práctico solo se exhibe desde la perspectiva del agente hacedor (Vega Encabo, 2009), esto se refiere a la valoración del conocimiento funcional y epistémico de los hacedores y el artefacto como portador de valor y normativa.

Además de diferenciar estos conceptos, debe tenerse en cuenta que algunas de las realizaciones pueden clasificarse como artefactos y como objetos, lo que no significa lo mismo, donde el artefacto surge o está relacionado con diversos estados mentales, fueron diseñados con un propósito o función intermediaria entre la propia intención en su creación y el artefacto en sí, en tanto que el objeto suele surgir a partir de efectos colaterales o accidentales de la acción humana (Vaccari, 2011). La función que distingue a los artefactos es el elemento central en la intención de su creación. La diferenciación entre artefacto y objeto indica que los objetos técnicos (artefactos) y los objetos estéticos (obras de arte) no son compatibles, aunque coincidan en materia, herramientas, forma, procesos de producción y resultado, la obra de arte es un objeto que tiene una función comunicativa, esto es: (1) un símbolo creado por el artista (significante); (2) un significado dentro de la conciencia colectiva; y (3) una relación entre quien la admira y su propia interpretación (Ré, 2014).

El estudio de la filosofía de los artefactos sucede en forma de proceso, el cual inicia por discurrir en la relevancia de las teorías científicas en la etapa de construcción de los artefactos, por tanto, nacen de la epistemología de sus hacedores (Mascoró, 2017). En este inicio del proceso, en la construcción de los artefactos, suceden distintas condiciones: (1) existen desde el momento de su diseño y planificación (delineación y esbozo, identificación de materiales, definición de la forma y condiciones sociales de utilidad del diseño); (2) se demanda una anticipación teórica donde se definen las intenciones de uso: condicionantes simbólicos, de habilidades, de acceso (social, económico o físico) y de interpretación; y (3) constitución del sentido y existencia del artefacto, representado en las formas como se alimenta, ya sea por energía o por información (Broncano, 2008).

Esta clase de procesos son considerados cadenas causales en los cuales se mide la eficacia de los artefactos, tomando como referencia al pen-

samiento contemporáneo según bases sociológicas, antropológicas y arqueológicas (Parente, 2016) y dan como resultado las siguientes tesis al respecto: (1) los artefactos no son intermediarios neutrales, ya que influyen en las percepciones; (2) el sentido ontológico que se propicia entre la parte humana y los objetos tecnológicos; y (3) las condiciones morales en las que influyen los artefactos a través de su intencionalidad y responsabilidad (Moreno, 2019). En este último punto, debe considerarse como aspecto pendiente de estudio la problemática filosófica que representan los organismos modificados genéticamente o las modificaciones corporales que incorporan artefactos al cuerpo humano, consideradas entidades complejas y que son producto de los artefactos creados con otra intención (Márquez, 2020).

De lo anterior, surgen diversas perspectivas psicológicas: (1) relaciones conceptuales y ontológicas entre función técnica y estructura física; (2) relación entre la función técnica, estructura física y uso previsto; (3) investigación epistemológica de la función de atribuciones; y (4) las condiciones sociales, funciones técnicas y su carácter intencional (Monterroza-Ríos, 2010). Como consecuencia, surgen diversos aportes fundados en las siguientes vertientes: (1) amplitud de trabajos experimentales monistas hacia la generación de bases pluralistas; (2) interacciones entre la vida experimental, el trabajo experimental y la cultura experimental; (3) estudios experimentales producidos especialmente en los laboratorios, basados en las teorías y contenidos representacionales; y (4) discusiones de problemas sobre la ciencia en cuanto a objetividad, evidencia, construcción de hechos y adecuación empírica (Moreno y Vinck, 2021). Los aspectos enumerados anteriormente resumen el proceso de las cadenas causales sobre la actividad investigativa a la que contribuyen los artefactos.

3. Un breve recorrido histórico de los aparatos epistémicos

A continuación, se describen diversos momentos históricos que resumen la presencia de los artefactos en la generación de conocimiento científico y en donde se pone de manifiesto que los procesos actuales relacionados con este aspecto de la filosofía de la ciencia y la tecnología se fueron transformando hasta llegar a la situación actual:

a) En el primer párrafo de su *Metafísica*, Aristóteles asienta que todos los hombres por naturaleza desean saber. Lo sustenta sobre la base de nuestro “amor a las sensaciones”. Amadas, — asegura— al margen de cualquier utilidad, y pone por encima de todas las otras sensaciones visuales (Aristóteles, ca. 350 a. C./2003, p. 69-70):

Y es que no solo en orden a la acción, sino cuando no vamos a actuar, preferimos la visión a todas — digámoslo — las demás. La razón estriba en que esta es, de las sensaciones, la que más nos hace conocer y muestra múltiples diferencias

De acuerdo al desarrollo que aquí se plantea, se asume que Aristóteles acierta en cuanto a que las sensaciones, principalmente las visuales, han mantenido una alta preponderancia al momento de comprobar el contenido veritativo de nuestras afirmaciones. Así, es de esperarse que los ejemplos que se ofrecerán a continuación descansan, en su mayor parte, en pruebas y demostraciones de corte visual.

b) En uno de los viajes a Egipto, Tales de Mileto debió plantearse cuál sería la altura de las pirámides, su ingenio entrenado tanto en las lides matemáticas y constructivas le llevaron a descubrir la hora exacta del día en que la sombra de su cuerpo tendría exactamente sus dimensiones reales, —se dice— así pudo calcular sin gran esfuerzo la altura de las pirámides egipcias (Laercio, 2007), las mismas que en su tiempo para su construcción requirieron de la invención y uso obligado de la denominada “megamáquina” (Mumford, 2010), es decir, la organización compleja del trabajo humano, si se tiene presente que un artefacto no es necesariamente material.

c) Otra historia de interés para los propósitos de este análisis es la de Eratóstenes y la medición de la circunferencia de la tierra, la cual hizo con una simple vara. Según comenta Guzmán Guerra (1999), mientras era director de la Biblioteca de Alejandría, Eratóstenes encontró un informe de observaciones efectuadas en Siena (actual Asuán), en Egipto. Dicho informe describía ciertas condiciones del medio día del 21 de junio, las cuales aprovechó para realizar cálculos muy aproximados a los que actualmente consideramos bastante exactos.

d) En la Edad Media, la cámara oscura fue un dispositivo que atrajo la atención de quienes estaban interesados en el fenómeno de la luz y se discute actualmente cuál fue su influencia en la pintura posterior al Renacimiento. Luego, pocos siglos después, se identifican otras referencias en Da Vinci (1498/1980), en su *Tratado de pintura*, donde estudia la proyección de cuerpos iluminados.

Aunque de ello tampoco se desprende ningún posible uso por parte de Da Vinci para sus creaciones artísticas, más bien, todo parece indicar, que se basó en ella a modo de una analogía en relación con el ojo, para tratar de explicar por qué, a diferencia de la cámara, no vemos las imágenes invertidas (Lindberg, 1976, 1968). La historia posterior de la cámara oscura consiste en una serie de mejoras dirigidas a dos fines: el perfeccionamiento de la

óptica y la invención de diferentes tipos de artefactos para tareas específicas. Probablemente, la necesidad técnica más urgente era la de poner al derecho la imagen, que se proyectaba al revés. Para ello se desarrollaron dos métodos principales, el más utilizado consistía en un espejo angulado para reflejar los rayos después de haber pasado por la lente (Newton, 1977; Kemp, 2000). Hockney (2001) ha reavivado una vieja polémica en torno a si los grandes artistas, aquellos que han realizado las obras de arte más logradas y relevantes utilizaron alguna clase de ayuda de tipo tecnológico. Esto le ha llevado a decir que esto produjo la primera revolución estética bajo la influencia de la tecnología, en particular de los artefactos ópticos utilizados durante los siglos XVII y XVIII.

e) El surgimiento del telescopio busca el acercamiento de la lejanía. El uso generalizado de los lentes, en múltiples usos cotidianos, fue creando una especie de furor por llevarlos a otros ámbitos. Ese es en parte el origen del telescopio, un invento que estaba ahí pero que nadie tuvo la ocurrencia u osadía de utilizarlo para examinar el espacio sideral, hasta que apareció Galileo. Su diseño amplió la capacidad de aumento al doble que la del telescopio anterior. Es decir, magnificó seis veces el tamaño de lo que se observaba. Poco después logra un telescopio que aumenta hasta veinte veces el tamaño de los objetos vistos.

Sus descubrimientos se dieron en medio de grandes polémicas, en principio porque se oponían decididamente a la concepción aristotélico/ptolemaica del mundo y por consiguiente a la autoridad de la Iglesia. Además, porque muchos de los filósofos naturales de la época, aunque partidarios de la observación como medio para obtener el conocimiento, ponían en tela de juicio la veracidad de las observaciones realizadas por medio de ese instrumento. Hoy tenemos telescopios fuera de la atmósfera terrestre enviando imágenes maravillosas a cada momento, deslumbrándonos no solo al mostrarnos espacios desconocidos, sino al mismo tiempo de la gran inventiva humana llevada a niveles simplemente insospechados.

f) El microscopio busca el engrandecimiento de lo más pequeño. Es Robert Hooke quien realiza las primeras incursiones en el mundo de lo inobservable con la mediación de un microscopio. Con base en una reelaboración de las influencias de Bacon y Descartes advierte la necesidad de reconocer los límites de la certeza matemática en la ciencia natural. Aunque Bacon nunca desarrolló una doctrina de los instrumentos de observación, solo menciona los microscopios y telescopios para mostrar su decepción de ellos. Hooke en cambio esbozó una psicología mecanicista de la

percepción y la memoria, además subrayó la importancia de la percepción en el proceso de inducción experimental, hasta el extremo de considerar que los sentidos poco fiables si no están “asistidos de instrumentos y armados de máquinas” (Solís, 1989), finalmente construyó él mismo dichos instrumentos. Desarrolló un microscopio compuesto utilizable y produjo la primera obra considerable dedicada en su totalidad a investigaciones microscópicas, la *Micrografía*, publicada en 1665. La habilidad de Hooke —piensa Solís— en este terreno quedaba de manifiesto por haber sido el primero capaz de construir microscopios aceptables para el trabajo científico y que se tomaron como modelo para los posteriores.

Otro de los personajes más destacados en la tarea de promover el cambio sísmico en el modo de ver el mundo que se produjo en el siglo XVII fue Antoni Van Leeuwenhoek. Aunque a diferencia de sus contemporáneos, que también utilizaban el microscopio, este autor se caracteriza por ver al microscopio como un instrumento óptico y

de aplicación práctica, tal como lo indica Snyder (2017) en *El ojo del observador*. Así se fue construyendo una sólida tradición intelectual y científica que nos develó una parte del mundo, oculta por largos siglos, gracias al uso de instrumentos para ampliar nuestras facultades cognitivas y perceptuales. Nos fue enseñando a ver u observar el mundo de un modo distinto, a cobrar conciencia de que más allá de lo que vemos, se esconde un amplio universo animado e inanimado.

g) El microscopio electrónico está basado en la teoría cuántica y su historia se concentra en tres momentos: (1) en 1925 la descripción de la naturaleza ondulatoria del electrón por parte de Louis de Broglie; (2) en 1927 Davison y Germer confirmaron experimentalmente la difracción de los electrones; (3) en 1932 Max Knoll y Ernst Ruska construyeron el primer microscopio de transmisión de electrones (TEM) que emplea electrones en lugar de luz; y (4) en 1932 se logra que los TEM aumenten un objeto hasta un millón de veces (Takeuchi, 2013).

Tipo de ítem	Asociación con los términos de la DGM	Alcance	Tipo de elemento	Insumo/producto informativo
Libros, folletos y pliegos impresos	Texto	Libros impresos en forma de monografía publicados por separado	Artefacto	Información objetiva (codificada)
Materiales cartográficos	Material cartográfico, diagrama	Mapas y planos en dos y tres dimensiones, cartas aeronáuticas de navegación y celestes, atlas, globos, diagramas de bloque, secciones, fotografías aéreas con fines cartográficos y vistas a ojo de pájaro (vistas de mapa)	Artefacto	Información subjetiva (por decodificar)
Manuscritos y colecciones de manuscritos	Manuscrito	Materiales mecanografiados o impresos como libros, tesis, cartas, discursos, etc., y textos legales manuscritos	Artefacto	Información objetiva (codificada)
Música	Música	Música impresa (en notas musicales)	Artefacto	Información subjetiva (por decodificador)
Grabaciones sonoras	Grabación sonora	Discos, cintas, rollos para piano y grabaciones sonoras en película (diferentes a las destinadas a acompañar imágenes visuales)	Artefacto	información subjetiva (sin codificar) intermediada por otro artefacto
Películas y videgrabaciones	Película, videgrabación	Películas cinematográficas y programas completos, compilaciones, avances, noticieros y noticieros filmicos, tomas de archivo y material inédito	Artefacto	información subjetiva (sin codificar) intermediada por otro artefacto
Materiales gráficos	Diapositiva, dibujo técnico, diorama, estampa, fotobanda, tarjeta de actividad, tarjeta mnemotécnica, original de arte, reproducción de arte, transparencia	Opacos (originales y reproducciones de arte tridimensional, diagramas, fotografías, dibujos técnicos) o elementos destinados a proyectarse o ser vistos (fotobandas, radiografías)	Artefacto u objeto	información subjetiva (sin codificar) con o sin la intermediación de un artefacto
Recursos electrónicos	Recurso electrónico	Recursos electrónicos que constan de datos (números, textos, gráficos, imágenes, mapas, imágenes en movimiento, música, sonido, etc.), programas (instrucciones que procesan datos para uso), combinación de datos y programas. Suceden en acceso directo (usando un dispositivo como disco, casete o cartucho, o un periférico) o remoto (computadora en red o disco duro o dispositivos de almacenamiento).	Artefacto	Información objetiva (codificada) sin intermediación de otro artefacto e información subjetiva (sin codificar) intermediada por otro artefacto
Artefactos tridimensionales y realia	Braille, juego, juguete, modelo, portaobjeto de microscopio, realia	Modelos, dioramas, juegos, casetes braille, esculturas y obras de arte tridimensionales, objetos de exposición, máquinas y vestuarios, objetos para ser exhibidos y aquellos que ocurren de forma natural o especímenes para microscopio	Objeto	Información subjetiva (por decodificador)
Microformas	Microformas	Toda clase de información en microformas (microfilms, microfichas, microopacos y tarjetas de ventana) de materiales impresos y gráficos	Artefacto	Información objetiva (codificada) sin intermediación de otro artefacto e información subjetiva (sin codificar) intermediada por otro artefacto
Recursos continuos	Texto	Publicaciones seriadas (finitas o infinitas), actualizaciones de hojas sustitibles, actualizaciones de sitios Web y boletines informativos de acontecimientos	Artefacto	Información objetiva (codificada)
Análisis	Texto	Describe parte o partes de un documento	Artefacto	Información objetiva (codificada)

Tabla I. Comparación entre tipos de ítems y características artefactuales

4. La filosofía de los artefactos y su relación con los estudios de la información: una breve aproximación

Estas cadenas causales de investigación científica usando artefactos, ponen de manifiesto la manera de vincular de forma transdisciplinar a las ciencias de la información con la filosofía de los artefactos, a través de dos perspectivas:

1. La relación con la bibliotecología tradicional, a través de su concepción de los objetos y los artefactos en las *Reglas de Catalogación Angloamericanas* (The Joint Steering Committee for Revision of AACR, 2004), que norman una de las bases fundamentales de la bibliotecología: los procesos de catalogación. Las RCA reconocen 12 tipos de ítems vinculados a la información, factibles de describir y de manera adicional, se incluye una lista propuesta de términos que facilitan o precisan el tipo ítem a través de una designación general del material o DGM (The Joint Steering Committee for Revision of AACR, 2004). En el intento por relacionar la filosofía de los artefactos con los tipos de ítems, según la Tabla I, se observa la siguiente caracterización: (1) un tipo de ítem puede funcionar como objeto/artefacto, uno como objeto y 10 como artefactos; y (2) el insumo/producto informativo observa las opciones de: Información objetiva (codificada) (cuatro tipos de ítems); información subjetiva (por decodificador) (tres ítems); información subjetiva (sin codificar) intermediada por otro artefacto (dos ítems); información subjetiva (sin codificar) con o sin la intervención de otro artefacto (un ítem); e información objetiva (codificada) sin intervención de otro artefacto e información subjetiva (sin codificar) intermediada por otro artefacto (dos ítems).
2. Las tendencias en perfiles profesionales vinculados con los estudios de la información. En la actualidad, los perfiles profesionales en ciencias de la información han superado a la bibliotecología tradicional, sin dejarla de lado, centrándose en aspectos relacionados con la producción científica y académica, especialmente en áreas como: estudios métricos de la información, comunicación y divulgación científica a través de diversos medios, arquitectura y sistemas de información, gestión del conocimiento y servicios de información, alfabetización informacional, análisis de datos y *machine learning* y web semántica, por mencionar los principales (Menéndez Echavarría et al., 2015; Arenas Grisales et al., 2022).

5. Los centros de investigación y sus artefactos como estudio del caso

La gravitación epistémica de dos microscopios electrónicos fundamentales para la generación de conocimiento, que es el tema central de este artículo, refiere a dos artefactos con los que trabaja el CIMAV (TEM y microscopio óptico) y que son motivo de estudio, ya que representan la forma actual de trabajo en la filosofía de los artefactos. Debe mencionarse que este centro público de investigación fue creado en 1994, con el fin de ofrecer asistencia tecnológica a la industria local; y en 2006, el Consejo Nacional de Ciencia y Tecnología (CONACYT) emitió una convocatoria para el “Establecimiento de Laboratorios Nacionales de Infraestructura Científica y Desarrollo Tecnológico (Biotecnología, Nanotecnología, Cómputo Avanzado, Energías Alternas)”, para equipar con infraestructura especializada a instituciones de ciencia, creándose así el Laboratorio Nacional de Nanotecnología *Nanotech* (Espinosa Magaña, 2016), el cual, está compuesto por dos instrumentos fundamentales en su equipamiento y funcionamiento:

1. El TEM de emisión de campo de ultra alta resolución *Nanotech TEM JEOL JEM-2200FS+CS* está “equipado con un cañón de emisión de campo (FEG) de 200kV y el filtro de energía en columna (filtro Omega), que permite obtener una imagen de pérdida cero (...) lo que da lugar a imágenes claras con alto contraste” (JEOL, 2023, párr. 1).
2. El TEM sirve para trabajar con materiales biológicos, denominado *Nanotech TEM Hitachi 7700*, tiene dos modos para la adquisición de datos de tomografía: modo bio y modo de alta resolución con resolución de punto optimizada, lo que permite que sirva tanto para las comunidades científicas que trabajan sobre los fenómenos de la vida, así como en la ciencia de materiales (Hitachi High Tech Corporation, 2023).

Ambos microscopios responden a los requerimientos de ser estudiados como artefactos y no como objetos, en consideración de lo siguiente: (1) funcionan como sistemas sociotécnicos a través de la retroalimentación en la interacción humana y artefactual; (2) los usuarios demandan de una epistemología suficiente, tanto disciplinar como investigativa; (3) permiten el dualismo artefactual entre función, intención o uso; (4) funcionan en relación a condiciones normativas definidas; (5) permiten el trabajo individual, sin embargo, la mayoría de sus actividades científicas se basan en el pluralismo; (6) se alimentan de información, caracterizada en ser subjetiva (por decodificar) como insumo, de lo cual se derivan productos científicos

de información objetiva (codificada), principalmente proyectadas a través de diversos elementos de producción y comunicación científica.

Además, con fines de demostrar los *affordances* de los artefactos estudiados (representados en la Figura 2), la justificación de ambos viene dada por la medición de su aprovechamiento práctico, a través del desarrollo de cadenas causales que buscan su rentabilidad o eficiencia, al realizar diversos actos de investigaciones técnicas formales con resultados objetivos, respondiendo a una cultura epistémica artefactual de acuerdo a las políticas científicas nacionales, los intereses económicos a través del desarrollo de proyectos con entidades externas a la institución, gestión de recursos financieros, humanos y materiales, y contribuciones en la formación y desarrollo de investigadores.

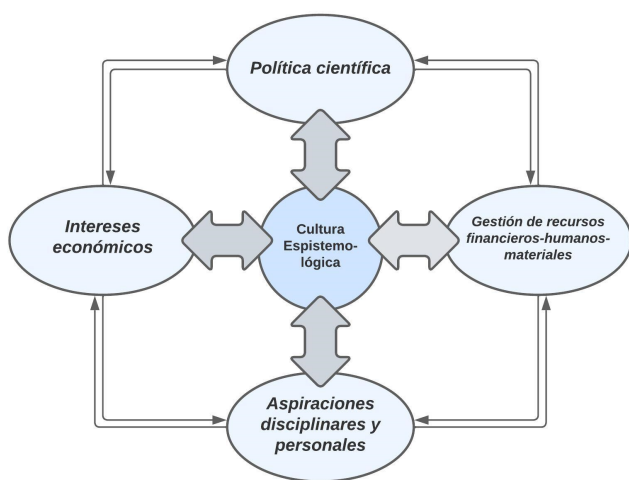


Figura 2. Elementos de cultura epistemológica en los affordances de los artefactos

De acuerdo a los datos recolectados, fue posible identificar que los dos artefactos mantienen acti-

vidades relacionadas con proyectos de investigación rentables, publicaciones científicas, formación de recursos humanos y formación de comunidades epistémicas, a través de 13 rubros de interés que responden a los elementos de la cultura epistemológica vigente, teniéndose hasta ahora como referencia de los *affordances* solo el registro de frecuencias de la actividad, sin una definición de un modelo de eficiencia del artefacto a partir del desarrollo mínimo aceptable de frecuencias de uso. Los rubros son: (1) proyectos (iniciativas financiadas con recursos externos: proyectos por convocatoria de CONACYT y proyectos financiados por el sector productivo); (2) tesis (número de muestras hechas con el artefacto para sustentar una tesis de programas académicos propios); (3) tesis externas (número de muestras hechas con el artefacto para sustentar una tesis con convenios interinstitucionales de México y América Latina); (4) convenio (colaboraciones interinstitucionales); (5) artículos (publicaciones científicas en revistas indizadas); (6) prototipo (desarrollo tecnológico para la solución de una problemática interna); (7) servicios importados (compra de servicios a otras instituciones); (8) estancias (estudiantes, personal técnico o investigadores que participan temporalmente en la propia institución o en otras); (9) vinculación T1 (acciones preestablecidas a través de un catálogo de servicios cobrados); (10) vinculación T2 (actividades científicas que requieren de análisis complementario donde una muestra pasa por al menos dos artefactos o dos pruebas); (11) vinculación T3 (actividades científicas que conllevan la interpretación de resultados por parte de un investigador, así como las propuestas de solución a las problemáticas que se identifican); (12) calibración (medición de la variación nanométrica); y (13) maquilado (fabricación de artefactos de apoyo a los procesos de investigación).

Renglones de uso del artefacto	Años de actividad (frecuencias de uso)										
	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	
Proyectos	299	360	457	442	318	93	96	34	32	107	
Tesis	543	775	639	531	346	348	284	113	277	391	
Tesis externas	103	193	174	112	137	135	194	60	20	12	
Convenios	227	243	103	120	127	46	16	0	2	13	
Artículos	401	426	373	331	176	303	229	171	323	298	
Servicios importados	34	53	0	0	0	0	0	0	0	0	
Estancias	0	74	245	60	50	222	271	174	127	168	
Vinculación T1	0	0	2	0	0	4	0	1	5	6	
Vinculación T2	0	0	6	5	0	2	0	0	4	0	
Vinculación T3	50	245	67	2	8	23	42	4	28	4	

Tabla II. Affordances del artefacto Nanotech TEM JEOL JEM 2200FS+CS

En relación con el artefacto Nanotech TEM JEOL JEM 2200FS+CS, este muestra actividad en 10 de los 13 rubros que componen el modelo de medición de efectividad durante el período 2013-2022, que de acuerdo a la Tabla II, se caracterizan por lograr siete productos de formación de comunidades epistémicas (proyectos, convenios, servicios importados, estancias y tres formas de vinculación), dos corresponden a la formación de recursos humanos (tesis y tesis externas) y uno relacionado con comunicación científica (artículos), estando así tres rubros ausentes (prototipos, calibración y maquilado).

El artefacto Nanotech TEM Hitachi 7700 observó el siguiente comportamiento: atiende seis de los 13 rubros enlistados en el modelo de efectividad de los artefactos. Los resultados observados en la Tabla III, correspondientes a los años de actividad del período 2016-2022, se refieren a productos formadores de comunidades epistémicas en tres renglones (proyecto, convenio y estancia), dos en formación de recursos humanos (tesis y tesis externas) y uno relacionado con comunicación científica (artículos). En este caso, se observa ausencia de siete rubros (prototipos, servicios importados, tres formas de vinculación, calibración y maquilado).

Tipo de muestra	Años de actividad						
	2016	2017	2018	2019	2020	2021	2022
Proyectos	50	29	169	50	24	32	47
Tesis	280	78	163	166	161	348	264
Tesis externas	56	44	50	134	62	5	25
Convenios	4	0	0	14	0	0	20
Artículos	117	54	236	141	79	259	224
Estancias	107	16	25	119	114	28	159

Tabla III. Affordances del artefacto Nanotech TEM Hitachi 7700

Las comunidades epistémicas se vuelven elementos axiales de la investigación científica y, por lo tanto, de la generación de conocimiento, la cual se está dando a partir de los microscopios como artefactos. A esto se agrega la necesidad de vincular este tipo de actividades con los estudios de la información para la lograr la objetivación del conocimiento a través de documentos científicos, mismos que exigen acciones vinculadas con la creación, divulgación y comunicación del conocimiento hecha desde los centros de investigación, además, desde las perspectivas formativas de los estudios de la información con perfiles profesionales competitivos y con capacidad de adaptación a situaciones cambiantes (Moreiro, 2010; Torres Vargas, 2019).

6. Conclusiones

El tránsito de este linaje instrumental ha propiciado una imbricación entre las diversas tradiciones culturales epistémicas, es decir, el conjunto de normas y cánones delineados por las comunidades científicas o tecnológicas donde se asientan los criterios que validan el conocimiento, con los linajes técnicos, esa estirpe de dispositivos o artefactos que se hermanan en coincidencias sobre sus características de diseño inherentes, de acuerdo a su función y materialidad. Los microscopios han mantenido desde su invención una línea evolutiva reconocible en cuanto a su funcionalidad, aunque exista una transformación material en su diseño. Los microscopios electrónicos tan apreciados en la investigación nanotecnológica son parte de una genealogía artefactual que nos permite pensar en una coevolución entre máquinas y tradiciones culturales. Esto ha terminado por hacer más difusas las fronteras entre lo natural y lo artificial. Una consecuencia, se piensa, derivada de ampliar posibilidades de intervención sobre el mundo de la materia, implica que todas estas geniales herramientas, además de ayudar a “ver” lo inobservable, también nos han llevado a su transformación, creando materiales que antes no existían en el mundo natural.

Hasta ahora, el análisis y registro de las posibilidades de acción de los artefactos se centra en la frecuencia de su uso en los laboratorios que forman parte de la estructura científica de un centro de investigación. Dada la ausencia en la identificación de un modelo específico de efectividad de los artefactos, aún falta una mayor claridad y profundización en torno a los alcances producidos por cada uno de ellos en áreas más específicas, además de buscarse una redefinición sobre los elementos de inserción de los estudios de la información en este tipo de contextos científicos. Por tanto, se esperaría que la tendencia investigativa sobre los temas de estudio de este artículo sucediera hacia la definición de condiciones vinculadas a la consolidación de acciones transdisciplinarias.

En cumplimiento a los objetivos planteados en esta investigación, estos se responden de la siguiente manera:

1. El análisis histórico sobre la evolución y propósitos de los artefactos asociados con la naturaleza de la luz, permitió justificar la influencia epistémica de los dos artefactos estudiados en esta investigación, considerando que su evolución científica actual, no sería posible sin su construcción y avance a través del tiempo. Debe indicarse que resulta complejo la disposición de artefactos generadores de conocimiento en organizaciones académicas

y científicas para el desarrollo de este tipo de estudios.

2. Experimentar la filosofía de los artefactos dentro del contexto de la filosofía de la ciencia, demanda una amplia identificación de fuentes de información, además, de la necesidad de sistematizar su conocimiento para que sea comprendido en ámbitos de las ciencias sociales y las ciencias exactas. Aunque la pretensión original podría referirse a la posibilidad de lograr una condición transdisciplinar del conocimiento filosófico con la rentabilidad de la ciencia y los estudios de la información, esto solo permitió llegar a establecer una condición de relación, con lo que se propiciará el desarrollo de futuras investigación para el logro de una relación disciplinar objetiva y formal.
3. El estudio del caso presentado, aunque complejo, simboliza una primera aproximación entre los centros de investigación, los estudios de la información y la filosofía de los artefactos, por lo que, tales acercamientos generan la posibilidad de estudios futuros, especialmente en consideración a la preocupación de la medición de la rentabilidad de los artefactos a través de diseños de modelos específicos, más allá del mero registro de su uso.

Referencias

- Arenas Grisales, Sandra Patricia; Giraldo Lopera, Martha Lucía; Ochoa Gutiérrez, Jaider; Tangarife Patiño, Ana María (2022). Posibilidad, riesgo e incertidumbre: análisis de tendencias en las ciencias de la información. // Revista Interamericana de Bibliotecología, ISSN: 2538-9866. 45:3, 1-25. <https://doi.org/10.17533/udea.rib.v45n3e347313>
- Aristóteles (2003). *Metafísica* (A. Rodríguez López, Trad.). Barcelona, España: Gredos (Obra original publicada ca. 350 a. C.).
- Asse, J (2015). Una teoría artefactual de los objetos matemáticos. Ciudad de México: UNAM.
- Broncano, F (2008). In media res: cultura material y artefactos. // *ArtefactoS*. ISSN: 1989-3612. 1:1, 18-32.
- Da Vinci, Leonardo (1980). *Tratado de Pintura* (A. González García, Trad.). Madrid, España: Editora Nacional (Obra original publicada ca. 1498).
- Guzmán Guerra, Antonio (1999). Introducción. En *Mitología del firmamento* (pp. 7-26). Madrid, España: Alianza.
- Espinosa Magaña, Francisco (2016). Laboratorio Nacional de Nanotecnología-CIMAV. // *Mundo Nano*. ISSN: 2007-5979. 9:16, 157-167. <https://doi.org/10.22201/ceiich.24485691e.2016.16.56910>
- Hitachi High Tech Corporation (2023). Hitachi inspire the next. <https://www.hitachi-hightech.com/global/en/> (2023-03-29).
- Hockney, David (2001). *El conocimiento secreto. El redescubrimiento de las técnicas perdidas de los grandes maestros*. Barcelona, España: Destino.
- JEOL (2023). *Direct Industry*. <https://bit.ly/3ZcOBzi> (2023-03-29).
- Kemp, Martin (2000). *La ciencia del arte: la óptica en el arte occidental de Brunelleschi a Seurat*. Madrid, España: Akal.
- Laercio, Diógenes (2007). *Vida y opiniones de los filósofos ilustres*. Madrid, España: Alianza.
- Lawler, Diego (2003). Las funciones técnicas de los artefactos y su encuentro con el constructivismo social en tecnología. // *Revista CTS*. ISSN: 1668-0030. 1, 27-71.
- Lee, Edward Ashford (2020). *The Coevolution: The Entwined Futures of Humans and Machines*. The MIT Press.
- Lindberg, David C (1968). The Theory of Pinhole Images from Antiquity to the Thirteenth Century. // *Archive for History of Exact Sciences*. 5:2, 154-176. <https://doi.org/10.1007/BF00327249>
- Lindberg, David C (1976). *Theories of vision from A-Kindi to Kepler*. Chicago, Ill.: University of Chicago.
- Márquez, Mónica María (2020). Artefactos y entidades naturales modificadas por medio de la biotecnología. // *Trilogía Ciencia Tecnología Sociedad*. ISSN: 2145-7778. 12:22, 197-215. <https://doi.org/10.22430/21457778.1242>
- Mascoró, Luciano (2017). La fase social del artefacto: la filosofía de la tecnología de A. Feenberg como complemento para el análisis Heideggeriano del útil. *Thémata. Revista de Filosofía*. ISSN: 2253-900X. 56, 127-149. <https://doi.org/10.12795/themata.2017.i56.6>
- Mejía Rendón, Sebastián (2023). ¿Hacia una teoría "unificada" de los artefactos técnicos? <https://bit.ly/3i34N7p> (2023-03-29).
- Menéndez Echavarría, Alfredo Luis; Quiñones Torres, Aída Julieta; Ordóñez Paz, Adriana Jedsabel; Herrera Soto, Liliana Margarita; Roza Higuera, Carolina; Cruz Mesa, Hernando; Melo González, Leonardo; Pérez Puerto, Yeni Magali (2015). Tendències investigadores de la ciència de la informació i la biblioteconomia a Iberoamèrica i al Caribe. // *BiD: textos universitaris de biblioteconomia i documentació*. ISSN: 1575-5886. 35, 1-12. <https://doi.org/10.1344/BiD2015.35.15>
- Monterroza-Ríos, Álvaro David (2010). Alcances y algunas críticas al programa Naturaleza Dual de los Artefactos Técnicos – DNTA. // *Trilogía. Revista de Ciencia, Tecnología y Sociedad*. ISSN: 2145-7778. 2, 55-66. <https://doi.org/10.22430/21457778.39>
- Monterroza-Ríos, Álvaro David (2011). Artefactos técnicos: ¿Cuál es el enfoque más adecuado? // *Estud.filos*. ISSN: 0121-3628. 44, 169-192.
- Monterroza-Ríos, Álvaro David (2019). El papel retroalimentador de los artefactos en el desarrollo de las técnicas humanas. // *Trilogía Ciencia Tecnología Sociedad*. ISSN: 2145-7778, 11:20, 49-65. <https://doi.org/10.22430/21457778.1286>
- Monterroza-Ríos, Álvaro David; Escobar, Jorge; Mejía Escobar, Jorge (2015). Por una revaloración de la filosofía de la técnica. Un argumento a favor del rol cultural de la técnica. // *Revista CTS*. ISSN: 1668-0030. 30:10, 265-275.
- Mitcham, Carl (1989). ¿Qué es la filosofía de la tecnología? Barcelona, España: Anthropos.
- Mitcham, Carl (1994). *Thinking through Technology. The Path between Engineering and Philosophy*. Chicago, Ill.: The University of Chicago Press.
- Moreiro, José-Antonio (2010). ¿Por qué 'Bologna' en los estudios de información y documentación? // *Anuario ThinkEPI*. 4: 41-43
- Moreno, Juan-Carlo (2019). Contribuciones al debate sobre la relevancia moral de los artefactos tecnológicos. // *Trilogía Ciencia Tecnología Sociedad*. ISSN: 2145-7778. 11:21, 91-117. <https://doi.org/10.22430/21457778.1327>
- Moreno, Juan-Carlo; Vinck, Dominique (2021). Encuentros entre filosofía de la ciencia, filosofía de la tecnología y

- CTS. // Trilogía Ciencia Tecnología Sociedad. ISSN: 2145-7778. 13:25, e1776. <https://doi.org/10.22430/21457778.1776>
- Mumford, Lewin (2010). El mito de la máquina. Logroño, Peñas de Calabaza.
- Newton, Isaac (1977). Óptica o tratado de las reflexiones, efracciones, inflexiones y colores de la luz. Madrid, España: Alfaguara.
- Parente, Diego (2008). La concepción Heideggeriana del artefacto en Grundbegriffe Der Metaphysik. // Signos Filosóficos. ISSN: 1665-1324. 10:20, 76-93.
- Parente, Diego (2016). Los artefactos en cuanto posibilitadores de acción. Problemas en torno a la noción de agencia material en el debate contemporáneo. // Revista Colombiana de Filosofía de la Ciencia. ISSN: 2463-1159. 16:33, 139-168.
- Parente, Diego (2020). El giro posthumanista en las humanidades y sus implicaciones para la filosofía de la técnica. Isegoría. // Revista de Filosofía Moral y Política. ISSN: 1130-2097. 63, 329-348. <https://doi.org/10.3989/isegoria.2020.063.03>
- Parente, Diego; Vaccari, Andrés (2022). Hacia una teoría de los artefactos como realizadores. // Contrastes: Revista Internacional de Filosofía. ISSN: 2659-921X. 27:3, 97-114. <https://doi.org/10.24310/Contrastescontrastes.v27i3.13975>
- Peterson, M. & Spahn, A (2011). Can Technological Artefacts Be Moral Agents? // Science and Engineering Ethics. ISSN: 1471-5546 17:3, 411-424. <http://doi.org/10.1007/s11948-010-9241-3>
- Quintanilla, Miguel Ángel (2020). ¿Por qué este libro es un clásico de filosofía de la técnica? // Kapp, Ernst (autor), Principios de una filosofía de la técnica. La génesis de la cultura desde un nuevo punto de vista (pp. 11-14). Oviedo, Asturias, España: KRK.
- Ré, Anahí Alejandra (2014). Tres modos de existencia del artefacto para pensar las poéticas tecnológicas. // Revista de Literatura, Historia e Memoria. ISSN: 1809-5313. 10:16, 09-23.
- Sandrone, Dario Rubén (2022). Objeto técnico. // Parente, Daniel; Berti, Agustín; Celis, Claudio (Coord.), Glosario de filosofía de la técnica (pp. 361-364). Buenos Aires: La Cebra.
- Snyder, Laura J (2017). El ojo del observador. Barcelona, España: Acanalado.
- Solís, Carlos (1989). Introducción. // Hooke, Robert (autor), Micrografía (pp. 11-76). Madrid, España: Alfaguara.
- Takeuchi, Noboru (2013). Nanociencia y nanotecnología. Ciudad de México: Fondo de Cultura Económica.
- The Joint Steering Committee for Revision of AACR (2004). Reglas de Catalogación Angloamericanas (2a ed.). Bogotá, Colombia: Rojas Eberhard.
- Torres Vargas, Georgina Araceli (2019). Formación para la investigación en bibliotecología y estudios de la información. // Biblioteca Universitaria. ISSN/e: 2594-0074. 22:1, 79-84. <https://doi.org/10.22201/dgb.0187750xp.2019.1.181>
- Vaccari, Andrés (2011). El artefacto, ¿estructura intencional o sistema autónomo? La ontología de la función artefactual a la luz del intencionalismo, el dualismo y la filosofía de Gilbert Simondon. // Revista Iberoamericana de Ciencia, Tecnología y Sociedad – CTS. ISSN: 1668-0030. 19:7, 197-208.
- Vega Encabo, Jesús (2009). Estado de la cuestión: filosofía de la tecnología. // Theoria. ISSN: 2954-4270. 66, 323-341.

Enviado: 2023-03-29. Segunda versión: 2023-06-15.
Aceptado: 2023-06-17.

Índice de autores

Author index

Escandell-Poveda, Raquel, 31
Iglesias-García, Mar, 31
Machin-Mastromatteo, Juan D., 51
Monteiro de Barros, Camila, 25
Morato, Jorge, 13
Moutinho, Sonia Oliveira Matos, 43

Papí-Gálvez, Natalia, 31
Ramírez Luján, Heriberto, 51
Sanchez-Cuadrado, Sonia, 13
Tarango, Javier, 51
Tarré Alonso, Beatriz, 25
Troitiño Rodríguez, Sonia Maria, 43

Índice de materias en español

Subject index in Spanish

Acceso a la información, 31
Artefacto epistémico, 51
Centro de Investigación en Materiales Avanzados
(Chihuahua, México), 51
Comunidades epistémicas, 51
Crítica diplomática, 43
Descripción de obras de arte, 25
Estudios bibliométricos, 43
Filosofía de la técnica, 51
Filosofía de la tecnología, 51
Filosofía de los artefactos, 51
Fragmentos enriquecidos, 13
Generación de conocimiento, 51
Google, 13

Maquinaria epistémica, 51
Mercado laboral, 31
Metadatos bibliográficos, 25
Metodologías digitales, 31
México, 51
Motores de búsqueda, 13
Optimización de respuestas, 13
Perfiles profesionales, 31
Práctica científica, 51
Producción científica, 43
Recogida de datos, 31
Recuperación de información, 25
Respuestas enriquecidas, 13
SEO, 13, 31

Índice de materias en inglés

Subject index in English

Access to information, 31
AEO, 13
Bibliographic metadata, 25
Bibliometric studies, 43
Centro de Investigación en Materiales Avanzados
(Chihuahua, Mexico), 51
Data acquisition, 31
Digital methodologies, 31
Diplomatic criticism, 43
Epistemic artifact, 51
Epistemic communities, 51
Epistemic machinery, 51
Featured snippets, 13
GAB, 13
Google, 13

Information retrieval, 25
Knowledge generation, 51
Labor market, 31
Mexico, 51
Philosophy of artifacts, 51
Philosophy of technique, 51
Philosophy of technology, 51
Professional profiles, 31
Rich answers, 13
Scientific practice, 51
Scientific production, 43
Search engine, 13
SEO, 13, 31
Web scraping, 31
Works of art description, 25