





Cultura y Big Data. Métodos  
y técnicas para el análisis cultural  
en una sociedad datificada

Cada libro de Colección Conocimiento es evaluado para su publicación mediante el sistema de dictaminación doble ciego por especialistas en la materia. Lo invitamos a ver el proceso de dictaminación de este libro transparentado en



[www.comunicacion-cientifica.com](http://www.comunicacion-cientifica.com)

[doi.org/10.52501/cc.014](https://doi.org/10.52501/cc.014)

Ediciones Comunicación Científica se especializa en la publicación de conocimiento científico en español e inglés en soporte del libro impreso y digital en las áreas de humanidades, ciencias sociales y ciencias exactas. Guía su criterio de publicación cumpliendo con las prácticas internacionales: dictaminación, comités y ética editorial, acceso abierto, medición del impacto de la publicación, difusión, distribución impresa y digital, transparencia editorial e indización internacional.

# Cultura y Big Data. Métodos y técnicas para el análisis cultural en una sociedad datificada

JESÚS EDUARDO OLIVA ABARCA



---

Oliva Abarca, Jesús Eduardo

Cultura y Big Data. Métodos y técnicas para el análisis cultural en una sociedad datificada / Jesús Eduardo Oliva Abarca. — Ciudad de México : Comunicación Científica, 2021. — 141 páginas. — (Colección Conocimiento).

ISBN: 978-607-99090-5-5

DOI: [doi.org/10.52501/cc.014](https://doi.org/10.52501/cc.014)

1. Big data – Aspectos sociales. 2. Tecnologías de la información – Aspectos sociales.  
3. Cultura. I. Título. II. Serie.

LC: QA76.9.B45

Dewey: 004.6

---

Primera edición en Ediciones Comunicación Científica, 2021



Diseño de portada: Francisco Zeledón

D.R. Ediciones Comunicación Científica S.A. de C.V., 2020

Av. Insurgentes Sur 1602, piso 4, suite 400,

Crédito Constructor, Benito Juárez, 03940, Ciudad de México, México,

Tel. (52) 55 5696-6541 • móvil: (52) 55 4516 2170

[infocomunicacioncientifica@gmail.com](mailto:infocomunicacioncientifica@gmail.com)

[www.comunicacion-cientifica.com](http://www.comunicacion-cientifica.com)

ISBN: 978-607-99090-5-5

DOI: [doi.org/10.52501/cc.014](https://doi.org/10.52501/cc.014)

Este libro es una publicación de acceso abierto con los principios de Creative Commons Attribution 4.0 International License que permite el uso, intercambio, adaptación, distribución y transmisión en cualquier medio o formato, siempre que dé el crédito apropiado al autor, origen y fuente del material gráfico. Si el uso del material gráfico excede el uso permitido por la normativa legal deberá obtener el permiso directamente del titular de los derechos de autor.

Esta publicación contó con el financiamiento otorgado por el Programa de Apoyo a la Investigación Científica y Tecnológica (PAICYT), y con el apoyo de la Facultad de Artes Visuales, de la Universidad Autónoma de Nuevo León.

# Contenido

<i>Agradecimientos</i> . . . . .	13
<i>Introducción</i> . . . . .	15
1. Micromecenazgo cultural y creativo en México y América Latina . . . . .	27
1.1. <i>Crowdfunding</i> , inteligencia colectiva y cultura de la participación . . . . .	27
1.2. <i>Web scraping</i> , o minería de red, para la recolección automatizada de datos . . . . .	35
1.3. Análisis de datos del micromecenazgo cultural y creativo . . . . .	47
2. El microrrelato digital: el caso de Twitter . . . . .	67
2.1. <i>Microblogging</i> y ficción en red . . . . .	67
2.2. Técnicas y aplicaciones del procesamiento de lenguaje natural . . . . .	75
2.3. PLN/NLP para el análisis y clasificación de microtextos digitales. . . . .	83
3. Visión computacional para el estudio de las artes plásticas . . . . .	97
3.1. Iconografía, iconología y teoría de la imagen . . . . .	97
3.2. Visión computacional: usos académicos e industriales . . . . .	104
3.3. Análisis de imágenes artísticas asistido por visión computacional. . . . .	110

*Conclusiones* . . . . . 127  
*Referencias* . . . . . 135

*A mi madre, Isabel, por su fortaleza,  
y a mi padre, Francisco, por su nobleza...*



*Technology can certainly be used to destroy both our physical world and our conceptual knowledge of it. But in the hands of inspired, creative and caring people, art and technology can also become means to convey empathy and build bridges between varied world views.*

KLAUS SCHWAB, Shaping the Fourth Industrial Revolution



## Agradecimientos

Este libro es, en gran medida, el resultado de diversos extravíos, de numerosos ensayos y errores, pero, también, de un grato proceso de aprendizaje que no ha estado exento de descubrimientos apasionantes, de conversaciones enriquecedoras y de debates sumamente provechosos. Por todo ello, el autor de este libro quisiera agradecer a muchas personas, pero en tales reconocimientos se podría escribir otro texto quizá de mayor extensión a la del presente; así pues, y en observancia de la brevedad y la concisión, quiero agradecer primeramente a la Universidad Autónoma de Nuevo León, dado que gran parte de este libro es el resultado del financiamiento otorgado a través de su Programa de Apoyo a la Investigación Científica y Tecnológica (PAICyT); también extendo mi gratitud a la Facultad de Artes Visuales, dependencia en la que me desempeño como profesor e investigador —siempre en formación—.

Agradezco, además, a la Facultad de Filosofía y Letras, lugar en donde ratifiqué, durante mi juventud, mi interés por la investigación y la academia, y a la Facultad de Artes Escénicas, en donde he tenido la oportunidad de entablar valiosos diálogos como profesor invitado. De manera muy especial, mi agradecimiento al Centro de Estudios Multidisciplinarios en Artes y Humanidades, espacio en el que mi disposición por investigar se tornó en una pasión; mi gratitud también al doctor Miguel Gabriel Ochoa Santos, a la doctora Elizabeth Sánchez Garay y al doctor Pol Popovic Karic, que son para mí figuras académicas inspiradoras, lo mismo que personas que tienen

todo mi aprecio y respeto. En este caso, el orden de aparición no es indicativo en absoluto del nivel de importancia del agradecimiento, por lo que deseo agradecer también a mis colegas, amigas y amigos de la Universidad Autónoma de Nuevo León, de la Universidad Autónoma de San Luis Potosí, de la Universidad Autónoma de Aguascalientes y de la Universidad de La Salle Bajío, con quienes he tenido la fortuna de compartir inquietudes y de intercambiar ideas.

Finalmente, mi gratitud también a mis estudiantes de licenciatura, así como a los de maestría y doctorado; a los primeros por su ingenio, su idealismo, su entusiasmo, su curiosidad, y su creatividad, cualidades que debieran acompañarnos en toda etapa de la vida; a los segundos, por sus ansias de conocer, el empeño y la pasión con la que llevan a cabo sus proyectos y por la resiliencia con la que, pese a toda duda e inconveniente, continúan su formación académica y personal. Todos ellos y ellas son un recordatorio permanente de que lo verdaderamente importante en toda travesía es el camino por recorrer. En el caso de este libro, reitero, el trayecto ha sido un tanto accidentado, pero tuve el acompañamiento de todas las personas que aquí se nombraron y aludieron; el arribo supone una gran satisfacción, pero es también un breve descanso para recuperar el aliento y continuar con un viaje que seguramente depara múltiples hallazgos.

## Introducción

Desde hace más de dos décadas, la civilización humana marcha a un paso cada vez más acelerado hacia la digitalización de las diversas dinámicas sociales. Las crecientes interdependencias e interconexiones económicas, políticas, tecnológicas y, también, culturales, que operan a escala global —globalización, para resumir—, así como la actual contingencia sanitaria que experimentamos como un mundo interconectado, han propiciado una incorporación todavía más veloz y extendida de las nuevas tecnologías de la información y la comunicación (TIC) en todos los ámbitos de la vida humana, así como una progresiva transición a un nuevo régimen socioeconómico que pudiera denominarse como sociedad “datificada” o sociedad de los datos, en la cual el dato se convierte en un recurso abundante para extraer lo mismo valor económico que conocimiento relevante sobre las interacciones sociales y las transformaciones y adaptaciones culturales en general.

En este tipo de sociedad perpetuamente vinculada mediante numerosos dispositivos —la computadora personal, la tableta, el *smartphone*, los *wearables*, etc.— generamos, intercambiamos y compartimos en todo instante datos relacionados con nuestras preferencias de consumo, nuestra filiación o aversión política, nuestros pasatiempos y nuestras actividades más cotidianas, hasta nuestra ubicación geográfica en tiempo real, junto con nuestros pensamientos más íntimos, nuestras quejas y opiniones respecto al curso que ha tomado el mundo —o al que, como humanidad, le hemos encau-

zado—. De igual manera, en este mundo datificado los pronósticos de la ciencia ficción nos son cada vez más familiares; el auge de la robótica, el internet de las cosas (IoT: *Internet of Things*), la inteligencia artificial y el aprendizaje máquina (*Machine Learning*), son tecnologías cada vez más cotidianas, al grado de que influyen también en nuestras acciones y en nuestras visiones de la realidad.

En este contexto, la ciencia, entendida en un sentido generalista como el análisis, la comprensión y la transferencia del conocimiento resultante del examen de los diversos aspectos que conforman la realidad (Olivé, 2007), enfrenta dos retos simultáneos de gran importancia: a raíz de los adelantos tecnológicos de nuestra época, el primer desafío es que la investigación científica requiere desarrollar una infraestructura que, apoyada en las nuevas tecnologías, posibilite la colaboración entre académicos e investigadores de múltiples disciplinas y ramas del saber, al tiempo que impulse la divulgación de su quehacer y la apropiación de sus aportes por parte de la sociedad (Hine, 2006); el segundo reto de la investigación científica en general consiste en desarrollar estrategias, técnicas y métodos tanto para recabar, estudiar e interpretar el *Big Data*, esto es, las cantidades abrumadoras de datos que generamos de manera constante y veloz (Manovich, 2020) y, al mismo tiempo, mantener vigente el diálogo histórico con los antecedentes teóricos que sustentan a las diferentes áreas del saber (Olivé, 2007).

Estos retos, que se vislumbraban desde finales del siglo xx, condujeron a la exploración de herramientas y técnicas asistidas por la informática y la computación como una alternativa para el abordaje de fenómenos cada vez más complejos —tanto cualitativa como cuantitativamente— que requieren de una visión inter-, multi- y transdisciplinaria, y de instrumentos que faciliten el manejo y tratamiento de datos a gran escala. Así, las aplicaciones de las ciencias computacionales y de las tecnologías de la información en la investigación científica en general derivan en la e-ciencia (*e-science*) (Gardner y Manduchi, 2007), a veces también denominada como *ciberciencia* (Hine, 2008), términos con los que se designa un paradigma epistemológico que exige reflexionar en torno tanto a la postura del investigador respecto a su quehacer (Estalella y Ardévol, 2011), así como a los alcances e influencia de la tecnología en la generación de conocimiento y su aplicabilidad para la mejora de la sociedad (Hine, 2006; Olivé, 2007) y, también, respec-

to a la brecha digital (Flew, 2014), esto es, las inequidades de acceso a tecnologías de apoyo para la investigación, sea por una cuestión de género o estrato social.

Es importante precisar que la e-ciencia, también denominada como e-investigación (*e-research*) (Anandarajan y Anandarajan, 2010), no se refiere únicamente a la incorporación de herramientas tecnológicas al trabajo científico, sino que consiste en el apoyo de la informática, el recurso a las bases de datos masivos y las redes colaborativas en línea en el ámbito académico (Estalella y Ardévol, 2011) para mejorar el flujo de trabajo de todo proyecto de investigación. Romero Frías insiste en que la e-investigación no se trata sólo de emplear herramientas computacionales para el análisis de datos, sino en una postura o aproximación al proceso de investigación en su totalidad, caracterizada por la automatización de ciertas etapas de la indagación (recolección y procesamiento de datos, contrastación de fuentes, etc.), la aceptación de fuentes de información y de divulgación no convencionales (blogs, redes sociales) y el manejo de enormes cantidades de datos (Romero Frías y Sánchez González, 2014, pp. 25-26).

La e-investigación es un planteamiento que fue primeramente empleado en las ciencias experimentales (Yang, Wang y Jie, 2011) y que, aunque enfrentó —o enfrenta todavía— ciertas resistencias, se ha incorporado paulatinamente a las ciencias sociales y a las humanidades (Burdick, 2012; Calderón, 2013), captando cada vez a más académicos, dada su dimensión transversal e interdisciplinaria (Estalella y Ardévol, 2011); de hecho, contrario a lo que pudiera pensarse sobre el término, la e-investigación no alude al adjetivo “electrónico”, sino al vocablo inglés *enhaced* (mejorado), lo que implica que se trata de una forma de investigar potenciada en y por recursos y tecnologías diversas (Romero Frías y Sánchez González, 2014). De acuerdo a esta sucinta caracterización de este paradigma epistemológico pudiera afirmarse que un proyecto de e-ciencia o e-investigación se formula principalmente en atención a dos factores: por una parte, a las características de los datos de los que se busca extraer información, esto es, su volumen, su complejidad y su velocidad de propagación; y, por otra parte, a los resultados esperados o aplicaciones que se busca obtener del estudio.

Así pues, existen proyectos de e-investigación que se apoyan fuertemente en el uso frecuente de recursos básicos del internet, tales como reposito-

rios digitales, bases de datos de publicaciones académicas, formularios en línea y otros servicios que brindan empresas como Google o Microsoft (Calderón, 2013; Chen, Wang y Cheung, 2010); otros trabajos de e-investigación recurren a herramientas más complejas de análisis de datos, incluso al empleo de lenguajes de programación, como Python o Java, por medio de los cuales se pueden realizar desde actividades de recolección de datos masivos mediante el acceso automatizado a servidores de bases de datos, o a través de las interfaces de programación de aplicaciones (API, *Application Programming Interface*) que brindan varios sitios web, hasta tareas de procesamiento, análisis y visualización de datos avanzadas (Gardner y Manduchi, 2007). Algunos proyectos de e-investigación implementan soluciones de cómputo en la nube (*Cloud Computing*) con miras a garantizar el trabajo colaborativo y la divulgación de los resultados de investigación (Terzo y Mossucca, 2017).

Como se sugirió anteriormente, las implicaciones que las TIC y el Big Data han tenido en todos los ámbitos de la vida humana no han pasado desapercibidos para las humanidades y las ciencias sociales, y desde hace años ya, investigadoras e investigadores de estas áreas emplean fuentes digitales y herramientas tecnológicas diversas para impulsar su quehacer. A raíz precisamente de tales prácticas potenciadas de investigación, surgen las humanidades digitales, campo de estudio en el que convergen tanto las aproximaciones teórico-metodológicas como los objetos de estudio de los estudios culturales, de la lingüística, de la teoría literaria y del arte, y de la sociología con la informática, concretamente con las técnicas y herramientas de recolección, extracción, procesado y análisis de datos de fuentes tanto análogas como digitales (Drucker, Kim, Salehian & Bushong, 2014). Las humanidades digitales se podrían considerar como una práctica de investigación “mejorada”, o de e-investigación, debido al uso de diversos *software*, así como de métodos computacionales, de ahí que también reciban el nombre de humanidades intensivas en *software* (*Software Intensive Humanities*) (Smithies, 2017).

En la actualidad —y desde sus orígenes— las humanidades digitales se consideran un ámbito académico de colaboración transdisciplinaria, integrando tanto a humanistas, científicos sociales, ingenieros computacionales, expertos en mercadotecnia, economistas, entre otros (Romero Frías y Sán-

chez González, 2014). Además, como indican Berry y Fagerjord, las humanidades digitales han adoptado, con singular entusiasmo, métodos y técnicas de la ciencia de datos y del análisis del Big Data, como lo son el análisis exploratorio de datos (EDA, *Exploratory Data Analysis*), el análisis de agrupamientos (*cluster analysis*) y las simulaciones computacionales. Otro rasgo en particular de las humanidades digitales es su interés creciente por el análisis de objetos culturales de origen específicamente digital, tales como *tweets*, *posts* en redes sociales, comentarios, imágenes y videos en diferentes sitios web (Berry y Fagerjord, 2017).

Dados los diversos enfoques que confluyen en las humanidades digitales, así como los variados objetos de estudio que admiten, se han generado aproximaciones teórico-metodológicas que parten de una disciplina en particular para una exploración más enfocada en cierto tipo de fenómenos en específico, como es el caso de la netnografía, o etnografía digital o virtual, que consiste en el análisis de las estrategias de autorrepresentación en comunidades en línea, así como de las interacciones de las personas en entornos digitales, como foros y redes sociales (Kozinets, 2010). Si bien la netnografía pudiera considerarse una vertiente de las humanidades digitales, entre sus especificidades se encuentran el que la primera se centra en el análisis de la vida *online*, mientras que las segundas, por su carácter comprensivo y su apertura, estudian también la gradual disolución de las fronteras que separan lo “real” de lo “virtual” —o *eversion*, término inglés que alude a la confluencia de lo interno con lo externo— (Jones, 2014).

La netnografía se enfoca, entonces, en lo que pudiera denominarse como cibercultura, esto es, las prácticas y modos de vida específicos de los internautas (Kozinets, 2010), así como las formas en que éstos se integran y erigen comunidades que se comunican de manera remota, a la vez que construyen o representan discursivamente sus identidades en el internet (Hine, 2015); las humanidades digitales, por su parte, no sólo estudian las formas de vida *online* y *offline*, analizan también artefactos u objetos —digitalizados o autóctonos de la cultura digital— que se clasifican como manifestaciones o concreciones de las actividades culturales humanas, como lo son los textos, las imágenes y los videos de todo tipo o género (Jones, 2014; Smithies, 2017). Conviene señalar, aunque parezca una obviedad, que la netnografía es una forma de investigación que se integra o que for-

maría parte del paradigma de la e-investigación al estar fuertemente centrada en las prácticas socioculturales que se efectúan en internet (Pink *et al.*, 2016).

Una de las formulaciones teórico-metodológicas más recientes en las humanidades digitales es la de la analítica cultural (*cultural analytics*), que, descrita de manera sucinta, integra y emplea de manera sistemática y explícita las técnicas de la ciencia de datos para el análisis y la visualización de macroconjuntos de datos culturales (Manovich, 2020). Tal aproximación se basa también en el flujo de trabajo y en las fases de desarrollo usuales de un proyecto de ciencia de datos (Peng y Matsui, 2016), que van desde la formulación del problema y/o necesidades a resolver, el minado de datos de diversas fuentes, como bases de datos abiertas, contenido textual y/o multimedia de sitios web y/o redes sociales (García Herrero *et al.*, 2018), hasta el análisis exploratorio, el modelado de datos y el despliegue de la aplicación web o móvil (Tandel, 2017). Este libro, que se propone como una práctica de e-investigación, se basa precisamente en la analítica cultural, propuesta formulada por Lev Manovich (2020).

Como todo objeto, acontecimiento o fenómeno inscrito en la sociedad datificada, los datos que pudieran denominarse como culturales se generan, se propagan y se comparten a una velocidad cada vez mayor y en cantidades más voluminosas: millones de personas —que cuenten con un dispositivo electrónico para acceder a internet— crean y comparten contenido digital diverso (ficciones, arte, videos, *podcasts*, etc.); ante esta situación, el estudio del Big Data cultural requeriría esfuerzos cada vez mayores simplemente para la recolección, la selección y el tratamiento de datos relevantes para una investigación; para esto, la analítica cultural parte de la premisa de que la alternativa más accesible para el análisis masivo de patrones, tendencias y dinámicas culturales es, precisamente, recurrir a los métodos y las técnicas de la ciencia de datos. Según este punto de partida, la analítica cultural se refiere entonces

al uso de métodos y diseños basados en las ciencias computacionales —incluyendo la visualización de datos, el diseño de medios interactivos, las estadísticas y el aprendizaje máquina— para la exploración y análisis de la cultura contemporánea a gran escala. Una de las metas de estas exploraciones es

permitirnos ver qué es lo que cientos de millones de personas alrededor del mundo crean, imaginan y valoran.<sup>1</sup> (Manovich, 2020, p. 9).

Conforme a la caracterización que de la analítica cultural realiza Manovich (2020), el flujo de trabajo de una investigación de este tipo comprende las siguientes fases: (a) la representación cuantitativa de datos referentes a un fenómeno cultural de interés, (b) la búsqueda o la generación de las fuentes para recabar los datos necesarios, (c) la organización de los datos en un conjunto coherente o corpus, (d) la exploración de los datos mediante visualizaciones o gráficos, (e) su análisis con métodos de la ciencia de datos, tales como estadística descriptiva e inferencial, modelización de los datos con técnicas de aprendizaje máquina supervisado y/o no supervisado y, de manera opcional, (f) el desarrollo, despliegue y difusión de aplicaciones interactivas para que toda persona —idealmente— tenga acceso tanto a los datos como a los resultados de la investigación (20).

Cabe señalar que la analítica cultural no es en absoluto un enfoque centrado en el aspecto cuantitativo de los datos; antes bien, su propósito es combinar los métodos y técnicas computacionales para el análisis de datos masivos con el estudio cualitativo de la cultura, optando así por un enfoque metodológico mixto, según el cual se generan premisas cualitativas —referentes a atributos o propiedades no medibles del objeto de estudio— a partir de la exploración de datos cuantitativos —del examen de las características ponderables del fenómeno abordado— (Creswell, 2014). A este respecto, el supuesto en el que se basa la analítica de la cultura consiste en que, de la identificación de patrones, tendencias, temas y conductas en macroconjuntos de datos culturales, es posible extraer conocimiento valioso relativo a cómo las personas construyen de manera cotidiana sentidos a través de sus interacciones, y cómo sus experiencias socioculturales están mediadas por lo digital (Manovich, 2020, p. 89).

Ahora bien, ¿qué es un dato cultural?, ¿qué clase de objetos o fenómenos se consideran “culturales” desde la óptica de la analítica cultural? Aunque

---

<sup>1</sup> En el original: “[...] to the use of computational and design methods—including data visualization, media and interaction design, statistics, and machine learning—for exploration and analysis of contemporary culture at scale. One goal of these explorations is to enable us to see what hundreds of millions of people around the world today create, imagine, and value”.

no se posea una definición “estable” del concepto, en este libro se asume que el término de dato cultural describe a un tipo particular de unidad de significación referente a manifestaciones o prácticas simbólicas específicas —la expresión creativa y/o estética de ideas, valores y cosmovisiones— creada, intercambiada, compartida y consumida de manera constante por millones de usuarios interconectados. Manovich (2020) señala que la analítica cultural aborda cuatro clases de datos culturales: el primer tipo alude a artefactos culturales tales como textos, obras de arte, fotos, música, diseño, películas, gráficos, arquitectura, juegos y aplicaciones de origen digital o, en su defecto, digitalizados.

El segundo tipo de datos culturales comprende información general sobre artistas u otros creadores culturales, direcciones de exposiciones artísticas, número de descargas de una aplicación, cantidad de veces que se ha “retuiteado” un tweet o compartido un estado o publicación en una red social, reseñas, puntajes, posts en los que los usuarios describen sus opiniones y experiencias —incluyendo, según sea el caso, fotografías y/o videos— respecto al trabajo de un artista o sobre una exposición, son también un tipo de dato cultural que corresponderían, en sociología del arte, a las elaboraciones ideológicas de segundo orden —textos de crítica, ensayos, entrevistas o demás trabajos sobre la obra— (García Canclini, 2014). La última clase de datos culturales que son objeto de la analítica cultural son las experiencias culturales, tales como el visionado de una obra teatral o de una ópera en algún servicio de *streaming*<sup>2</sup> (retransmisión o transmisión en directo) o jugar un videojuego (García Canclini, 2014, pp. 76-77).

Tras este vistazo panorámico al paradigma de la e-investigación, a las humanidades digitales como planteamiento potenciado para el estudio de los fenómenos culturales en general, así como a los enfoques específicos que se derivan de éstas, como son la netnografía y la analítica cultural, conviene ahora profundizar en algunos aspectos relevantes de este libro. Primero, en su conjunto, este texto constituye una tentativa de desarrollar un proyecto relativamente ambicioso de e-investigación mediante el planteamiento de la analítica cultural tal cual ha sido formulada por Lev Manovich; este libro es el resultado de un proyecto de investigación general integrado a su vez

<sup>2</sup> Véase Luzardo, A., Funes, G., & 23 Design. (2019). *Emprendimientos tecnocreativos: Creatividad y tecnología, ¿aliados o enemigos?* Inter-American Development Bank. <https://doi.org/10.18235/0001854>.

por tres proyectos desarrollados a partir de mediados del 2018 y concluidos a finales del 2020, principalmente por las necesidades particulares de actualización de las librerías o módulos programáticos y de los datos a partir de los cuales se elabora cada estudio.

Segundo, de manera paralela al desarrollo simultáneo de cada uno de estos proyectos, y aunque el autor ya poseía alguna experiencia en lógica programática y en codificación con los lenguajes de programación Python y R, la complejidad y las características específicas de cada uno de los estudios comprendidos en este libro requirieron que su autor aprendiera e incorporara técnicas e implementaciones de lógica computacional y de la ciencia de datos al estudio de los fenómenos culturales que se abordan en cada proyecto aquí presentado; a la par de este aprendizaje, difícil, pero también muy satisfactorio, una de las principales preocupaciones del autor, y razón de múltiples titubeos, era relativa a los aspectos epistemológicos implicados en un macroproyecto de este tipo: ¿es posible utilizar la programación y la ciencia de datos en beneficio del estudio de la cultura y las artes, de la economía creativa en general?, ¿o supone esta aproximación una forma de reduccionismo de la singularidad y complejidad de los fenómenos culturales?

Afortunadamente, la propuesta de Lev Manovich fue oportuna para disipar estas dudas y tener la certeza de que el abordaje de la cultura, a partir de técnicas y métodos de la ciencia de datos, puede enriquecer sustancialmente el conocimiento de las prácticas y manifestaciones culturales y creativas, así como posibilitar la transferencia y aplicabilidad del saber en beneficio tanto de la sociedad como de los agentes culturales mismos —artistas de cualquier disciplina, o inter-, multi- o transdisciplinarios, gestores y promotores culturales y demás profesionales creativos—. Así pues, y con la intención de que cada uno de los estudios aquí desarrollados brinden algún tipo de beneficio al sector cultural y creativo, se presentan también, como resultado de cada proyecto, tres aplicaciones web accesibles a todo usuario o usuaria, en las que podrá examinar los conjuntos de datos con los que se desarrolló cada estudio y también visualizar los resultados obtenidos que se describen, aunque panorámicamente, en el capítulo correspondiente.

Tercero, el autor ha apostado por dos criterios de estilo académico que, espera, enriquezcan la comprensión del lector o lectora o, en su defecto, no la menoscaben. Se ha optado, primero, por la concisión y la brevedad, in-

tentando así mantener un lenguaje que conserve cierta formalidad —acorde con la retórica científica esperada de un libro de investigación— y que a la vez comunique de manera clara y sin rodeos innecesarios las formulaciones teóricas, los métodos y técnicas, y los hallazgos que conforman cada capítulo del libro. Segundo, en la medida de lo posible, el autor invita —tal vez de manera insistente— a los lectores a que reproduzcan ellos mismos, con apoyo de las aplicaciones web compartidas en cada capítulo, los análisis que se llevaron a cabo en cada estudio, con la intención de que se formulen ellos o ellas sus propias reflexiones y conclusiones.

Con respecto a los estudios que comprende el libro, en el primer capítulo se presenta el desarrollo de un proyecto de investigación centrado en el análisis del *crowdfunding* —o micromecenazgo— de actividades artísticas, culturales y creativas. En la primera sección se describe en qué consiste el fenómeno, así como sus particularidades en lo que concierne al ámbito de la cultura y de la creatividad. En la segunda parte del capítulo se detallan las técnicas de recolección y preprocesamiento de datos, a la vez que se introducen conceptos relativos a las tareas y operaciones de adquisición automatizada de información; en la sección final se presentan los resultados del análisis del micromecenazgo cultural y creativo, y se muestran también las funciones generales de la aplicación web desarrollada durante esta investigación.

En el segundo capítulo, la primera sección aborda los temas de la textualidad en internet, los géneros textuales y las plataformas de divulgación de textos, específicamente de los microtextos de ficción y los sitios o redes sociales de *microblogging*; se examinan también algunos planteamientos teóricos relativos al microrrelato literario y a los problemas en torno al estudio de la ficcionalidad. En la segunda sección se examina el tema del procesamiento de lenguaje natural (NLP, *Natural Language Processing*) y sus aplicaciones en la vida cotidiana, lo mismo que en la industria y en la academia; aquí mismo se detallan los pormenores sobre la recolección de datos textuales y la conformación de un corpus compuesto por tres clases de textos diferentes; el capítulo concluye con la aplicación de algunas de las técnicas de procesamiento de lenguaje natural para la indagación de características o propiedades que diferencien a la microficción de otros tipos de microtextos; finalmente, se expone una aplicación

web que incluye un sistema automatizado de clasificación textual basado en el corpus analizado.

El último capítulo inicia con una breve revisión de los planteamientos y enfoques teóricos para el estudio de la imagen en general, así como de la imagen artística o, en particular, de obras de arte, ahondando en la semiótica visual y en el método iconográfico-iconológico; en el segundo apartado se abordan los empleos cotidianos de la visión computacional, así como sus aplicaciones en diferentes sectores productivos, a la vez que se explican las técnicas y operaciones básicas de procesamiento automatizado de imágenes. De igual manera que en los capítulos anteriores, en la sección final se presenta una aplicación desarrollada durante este estudio, cuyo propósito es demostrar los usos posibles de la visión computacional —así como sus ventajas y limitaciones— para el análisis de imágenes de obras artísticas.

Antes de cerrar esta introducción, es pertinente agregar que este proyecto general de investigación, que compendia, como ya se ha mencionado, tres estudios diferentes, es también resultado de una serie de motivaciones específicas. El primer y fundamental motivo es un genuino interés del autor por comprender las complejas dinámicas operantes en las artes, la cultura y la creatividad o, si se prefiere, en la economía creativa —su estructura organizativa, las actividades que la conforman, sus interacciones con otros sectores de la sociedad, etc.—; la segunda motivación es de un carácter más “utilitarista”, y consiste en el afán de averiguar cómo y mediante qué estrategias y herramientas se pueden mejorar las condiciones de producción, circulación y consumo cultural, en otras palabras, cómo se puede contribuir, desde la academia, al desarrollo de las artes, la cultura y la creatividad, tanto en sus dimensiones estéticas e intelectuales, lo mismo que en lo que respecta a su economía.

Aunado a lo anterior, existe también un interés por emplear la analítica cultural como un planteamiento de apoyo para artistas, creadores en general, promotores, gestores y demás agentes culturales y profesionales creativos en su incursión y/o desempeño en un ámbito que afronta innumerables desafíos. Además, dada la situación actual, que como sociedad global vivimos, es pertinente agregar un cuarto motivo que consiste en proveer al sector cultural y creativo, aunque sea al nacional, de una serie de herramientas y recomendaciones por medio de las cuales puedan subsistir tanto los

creadores como los espacios culturales en una época en la que una contingencia sanitaria planetaria obliga al mundo a reducir la interacción social en espacios concurridos y a adoptar el distanciamiento como medida preventiva ante la pandemia. En este contexto, las artes han tenido que asumir la digitalización como una alternativa para su difusión, sin embargo, no todos los involucrados en el sector de las industrias culturales y creativas cuentan con el conocimiento o con la infraestructura para lograr una presencia digital. En este sentido, este proyecto, y cada uno de los estudios que lo integran, pretende brindar a artistas, gestores, promotores y profesionales creativos en general, opciones e ideas para la transición a entornos digitales.

# 1. Micromecenazgo cultural y creativo en México y América Latina

## 1.1. Crowdfunding, inteligencia colectiva y cultura de la participación

El concepto de “campo”, en la acepción empleada por Bourdieu (2002), se define como un espacio de interacciones objetivas, organizado según las posiciones que los agentes involucrados ocupan al interior de éste. En el caso del campo de producción cultural, la legitimación de una práctica expresiva o de una forma artística particular depende de que se efectúen ciertas dinámicas propias o específicas del campo, que se llevan a cabo por medio de las interacciones entre creadores, intermediarios y consumidores, entre los que existen ciertos “pactos” —implícitos o más o menos explícitos—, que van desde la etiqueta o comportamiento en un museo o en un evento escénico hasta la adhesión a una estética o una preceptiva creativa. Todo ello deriva en la distinción entre prácticas culturales “consagradas” —que requieren de saberes más especializados y de la adopción de actitudes diferenciadoras— y “populares” —aquellas usualmente relacionadas con el entretenimiento—.

Un ejemplo contemporáneo de esta cuestión lo constituye el coleccionismo de ejemplares literarios inusuales o de piezas pictóricas o escultóricas originales al lado del consumo de series televisivas populares; en el primer caso, el consumidor se sitúa en una posición de mayor jerarquía en el campo de producción cultural, como un “iniciado” de las artes, mientras que en el segundo, la audiencia no está sujeta al cumplimiento de los requisitos que al conocedor de las artes, al individuo “de buen gusto”, se le exigen (Bourdieu, 2002). La organización por posicionamientos que configura al campo

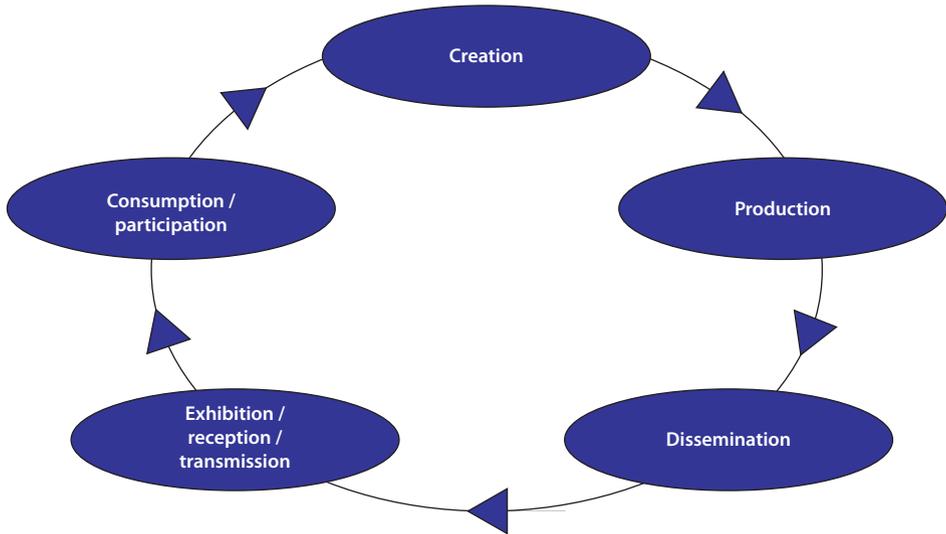
de producción cultural, según la óptica del sociólogo francés, permitiría explorar de manera objetiva las interacciones entre artista, obra y público, así como analizar el gusto o la percepción de las audiencias respecto a las manifestaciones artísticas.

Tales dinámicas, que competen tanto a la producción como a la circulación y al consumo de bienes, prácticas y manifestaciones que se catalogan como culturales, son, en cierta medida, resultado de una serie diversa de factores no sólo de índole social, sino también política, tecnológica y económica. En este sentido se puede asumir como válida la premisa de que el arte, y la cultura en general, son, más que la representación de las relaciones sociales de una época o de un espacio geopolítico, la concreción o la realización de los modos de figuración, de apreciación, de valoración y, principalmente, de organización de una sociedad en particular (García Canclini, 2014); no es, pues, inexacto afirmar que, como aseguran diversos autores y autoras, la cultura es portadora del sentido de lo humano, de su capacidad creadora y de su historicidad (Bauman, 2002; Cuche, 2002; McRobbie, 2016; Towse, 2013).

Dado que las dinámicas del campo cultural son partícipes de las transformaciones y acontecimientos históricos, cabe preguntarse cómo se modifican o, en su caso, se renuevan los procesos de la creación, la producción, la difusión, transmisión y consumo —o participación— de la cultura en una sociedad estructurada por las interconexiones que posibilitan las redes de información y comunicación digital o, en otras palabras, cómo se transforma el ciclo de producción cultural (ver figura 1) en una sociedad interconectada, fuertemente basada en la diseminación —cada vez más veloz e instantánea— de la información (Castells, 2000); con toda seguridad, la creación artística, y toda otra forma de producción cultural o práctica creativa, está estrechamente ligada a las tecnologías de la información y la comunicación (TIC) que sustentan a la sociedad actual (Goriunova, 2012).

Así pues, los mundos del arte, que conforman el campo cultural según la acepción que se ha elaborado aquí, se transforman, se adaptan, reflejan y también se resisten a una cotidianeidad digitalizada de múltiples e incesantes conexiones. Cabe señalar que el término “mundos del arte” se emplea de manera complementaria al de campo de producción cultural para aludir a las diferentes formas de organización operantes en cada sector cultural y

Figura 1. Ciclo cultural. UNESCO



creativo —o expresión artística, si se prefiere—, así como a los modos en que se distribuyen las actividades al interior de cada sistema —la “división” del trabajo artístico, y creativo, junto con los procesos de distribución y consumo específicos— (Becker, 2008; Peist, 2012). Si se parte, pues, de la premisa de que existen vínculos innegables entre la cultura y la estructura general de la sociedad, así como de que las tecnologías informacionales y comunicativas actuales ejercen un notable influjo en las prácticas culturales, se perfila como necesario el análisis de las formas de producción simbólica en los entornos digitales interconectados.

En el caso de este estudio, el fenómeno que se aborda es el del micromecenazgo de proyectos culturales y creativos en internet, práctica también denominada fondeo colectivo, o crowdfunding, que consiste en una forma de financiamiento colaborativo que opera mediante la promoción de proyectos en redes sociales y sitios web dedicados a reunir las aportaciones de patrocinadores a cambio de contraprestaciones que van desde recompensas —como en el caso de Kickstarter, Indiegogo e Ideame—, participación como inversor —Play Business, por ejemplo—, préstamo —como Kubo Financiero y Prestadero— y donación —ejemplos de esta modalidad son GoFundMe y Donadora—. Los proyectos que se promueven en estas plataformas van

desde obras y eventos artísticos y trabajos creativos diversos hasta programas de intervención social e innovaciones científicas y tecnológicas.

Al ser un modelo de apoyo económico basado en colaboraciones de numerosos patrocinadores, el micromecenazgo formaría parte de lo que Jenkins denomina como cultura de la participación, en la que las audiencias, o el consumidor, intervienen de manera más directa en los contenidos que circulan a través de los medios de comunicación masiva (Jenkins, 2008), principalmente en el internet. Esta forma de producción cultural participativa propicia nuevas interacciones entre creadores, empresas e instituciones, y consumidores, adquiriendo estos últimos un papel central en lo que se podría considerar una forma de “empoderamiento” del fan (Jenkins, 2006), lo mismo que en un nivel de *engagement* mayor al del aficionado ocasional (Mittell, 2013), que se convierte en una especie de co-creador o, si se prefiere, co-gestor y co-promotor, en tanto que no solamente participa de la expansión mediática de una obra o de una franquicia, sino que también la promociona e incluso puede llegar a tener injerencia en las decisiones creativas y comerciales del proyecto.

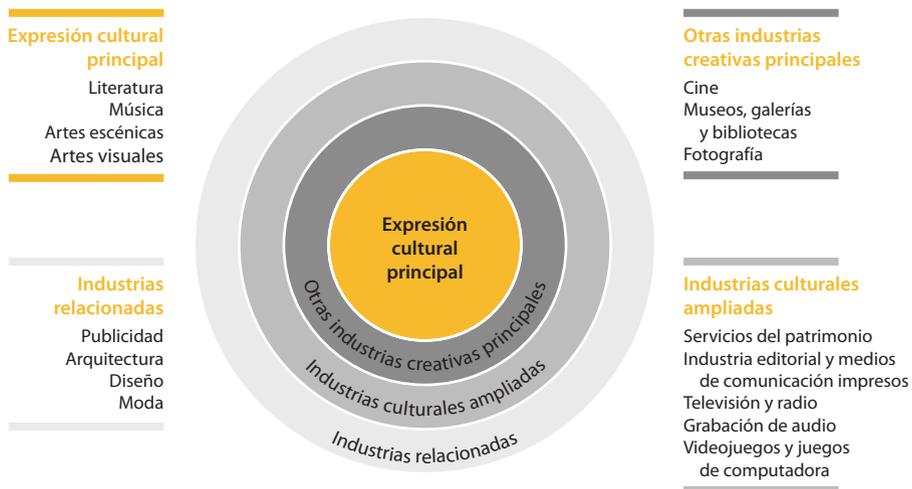
El micromecenazgo se vincula igualmente con el concepto de inteligencia colectiva, que podría definirse como una forma de cooperación consciente entre individuos pertenecientes a una misma comunidad o grupo para la realización de una tarea, la solución de una problemática o la obtención de un beneficio que les es de interés común (Flew, 2014). Si bien, el concepto tiene sus bases en el estudio sociológico y antropológico de la organización y evolución del comportamiento grupal, se utiliza actualmente para explicar las colaboraciones remotas, síncronas o asíncronas, que mediante la comunicación digital conllevan a la formación de comunidades en línea que los mismos miembros gestionan y administran, compartiendo así sus saberes en pro del grupo (Lévy, 2013). En concordancia con esto, otra de las características distintivas del micromecenazgo es el de conformar comunidades en las que, según la escala de la aportación, cada colaborador interviene activamente en las diferentes fases de desarrollo de un proyecto.

En lo que respecta a la microfinanciación de proyectos culturales y creativos, que es el objeto de este estudio, es pertinente acotar aquí dos aspectos teóricos relevantes: el primero es que las acepciones de los términos “cultural” y “creativo” se corresponden al concepto de industrias creativas, vocablo

que engloba en un mismo sector actividades que, si bien difieren respecto a sus finalidades o productos, se basan de igual manera en la creatividad y en la innovación —artística, intelectual o estética— (Stoneman, 2010) para el desarrollo de nuevos contenidos lo mismo que para nuevos modelos de negocios (Hartley, Wen y Siling Li, 2015). Es conveniente mencionar, además, que la noción de industrias creativas suprime la distinción entre una alta cultura y una cultura para las masas, o entre artes “populares” y artes “cultas” (Towse, 2013); en otras palabras, los proyectos culturales y creativos que se analizan aquí remiten de manera indistinta tanto a formas culturales “consagradas” (como la literatura, el teatro y la pintura) como a “populares” (cómic y videojuegos, por mencionar algunas).

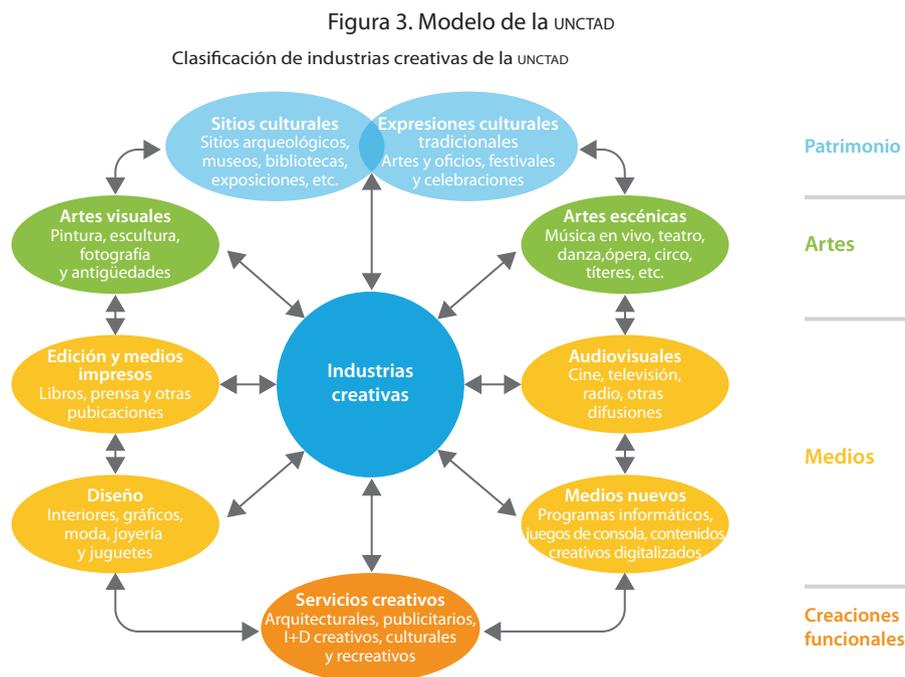
Segundo, existen diversos modelos de clasificación de las industrias culturales y creativas, de entre los que destacan el de los círculos concéntricos y su variante expandida (ver figura 2), así como el modelo de la Organización Mundial de la Propiedad Intelectual (OMPI) y el desarrollado por la National Endowment for Science, Technology and the Arts (NESTA), en Reino Unido, denominado como “modelo refinado de las industrias creativas”, además de la United Nations Conference on Trade and Development (UNCTAD), que parte de la distinción de cuatro subsectores clave: el de las artes como tales, en el que se incluyen las artes visuales y las escénicas, el de

Figura 2. Modelo de los círculos concéntricos. PNUD (2014)



los medios, que reúne a la industria editorial y a la audiovisual, el del patrimonio, que comprende los sitios históricos, los museos, y las celebraciones tradicionales, y el de las creaciones funcionales, que se refiere al diseño, a la arquitectura y a la publicidad, entre otras (PNUD, 2014).

Es necesario señalar que cada una de las tipologías antes referidas remite a una serie de problemáticas importantes para la reflexión en torno a los intercambios entre cultura, economía, política y tecnología, lo mismo que sobre el consumo y la valoración de la producción cultural y creativa en general; no obstante, y por las características de este análisis, tal cuestión, si bien de gran relevancia, no será tratada aquí. Así pues, para la conformación de la muestra para el estudio del *crowdfundig* cultural y creativo se siguió el modelo de la UNCTAD (ver figura 3), el cual reúne de manera no jerárquica actividades tales como la creación plástica (pintura, escultura), la producción multimedia (fotografía, cine, video, editorial, cómic, juegos), los espectáculos escénicos (música, danza, teatro), trabajos funcionales (diseño gráfico) y artesanales.



Aclaradas estas dos cuestiones, este estudio se propone explorar las dinámicas de producción cultural en México y en América Latina a través de los patrones, tendencias y anomalías que se identifiquen mediante el análisis de los proyectos culturales y creativos alojados en sitios de microfinanciación colectiva; los objetivos de esta investigación son, primero, bosquejar un panorama general de los procesos de creación, circulación y consumo —o participación, en este caso— simbólicos a partir de un fenómeno que se ha popularizado y extendido en toda Latinoamérica y, segundo, se espera que de los resultados de este análisis se deriven hallazgos de utilidad para las actividades de creadores, promotores, gestores y demás agentes que intervienen en el ciclo de producción cultural.

Uno de los aspectos de mayor interés del micromecenazgo cultural y creativo radica en el hecho de que evidencia la problemática, cada vez más extendida en los diferentes sistemas artísticos, de la financiación de las actividades creativas. Desde un punto de vista histórico, las plataformas de microfinanciamiento colectivo serían la alternativa a la escasez de individuos o particulares acaudalados, aficionados al arte y con una actitud usualmente diletante, que eran conocidos como mecenas, así como de los patrones, que subvencionaban el trabajo artístico con fines políticos (Heinich, 2002; Maanen, 2009). Ante la paulatina desaparición del mecenazgo y del patronaje individual, el Estado asume la responsabilidad de sustentar el arte, hecho que, desde cierto punto de vista, implica un límite a la libertad artística y a la realización de valores estéticos intrínsecos en pro de una determinada ideología política (Maanen, 2009).

De manera paralela, y con un manifiesto interés comercial, empresas y organizaciones con fines de lucro patrocinan también las actividades artísticas, aunque a este respecto es necesario aclarar que existe una empresariedad cultural, esto es, centrada en la comercialización de bienes y servicios simbólicos (Rowan, 2010) y una empresariedad que se apoya en la cultura como un medio. Como señala acertadamente Zallo (1992), la cultura siempre ha sido apropiable y, por lo tanto, la “empresariedad” cultural ya existía —en menor escala— desde los siglos XVIII y XIX, durante los cuales la literatura, principalmente la novela, y la pintura conforman un mercado importante —las editoriales, las librerías, las galerías de arte, junto con las actividades de los *marchands*, son evidencia de ello— (Becker,

2008; Poli, 1976). No obstante, la cultura es también un recurso para la consecución de otros fines distintos a los culturales (Yúdice, 2002; Zallo, 1992), tales como la condonación de impuestos o el posicionamiento de una marca.

Dado que los apoyos y los recursos tanto gubernamentales como los de la iniciativa privada son limitados y están sujetos a convocatorias y/o concursos, otra alternativa para la financiación del trabajo artístico —y cultural y creativo en general— es la del emprendimiento y el autoempleo, que se conciben usualmente como una forma de reivindicación y liberación del artista (Gander, 2017) de cara a regímenes laborales monótonos, impuestos por empresas despersonalizadas, así como a la necesidad de recurrir a otros empleos para financiar el trabajo creativo que, a su vez, se marginaliza a las horas libres o de ocio (Throsby, 2008); cabe señalar que tal versión del emprendimiento y del autoempleo también puede considerarse como una panacea para el desempleo, la precariedad y el condicionamiento de la inversión pública y privada para el desarrollo del sector cultural (McRobbie, 2016; Rowan, 2010).

El micromecenazgo cultural y creativo se presenta entonces como una alternativa a la limitación de recursos y fondos gubernamentales y privados, así como a las problemáticas de la restricción de la autonomía creadora y de la subordinación de valores estéticos a fines políticos o comerciales, además de potenciar iniciativas de emprendimiento a través de comunidades de usuarios que no solamente muestren interés en los proyectos como productos finalizados, sino que participan en todo el desarrollo de los mismos e incluso contribuyen, tácita o manifiestamente, a la comercialización de éstos. No obstante, es válido preguntar si los proyectos culturales y creativos financiados de manera colectiva en internet no están sujetos a limitantes específicas de la cultura digital como, por ejemplo, el imperativo de la novedad (Yproductions, 2009) y la creación centrada en los gustos de la audiencia antes que en la realización de una visión creativa libre (Hesmondhalgh, 2013; Maanen, 2009; McRobbie, 2016; Throsby, 2008), la brecha o inequidad en el acceso a los medios digitales (Flew, 2014) y la necesidad de construir y gestionar redes colaborativas (Goriunova, 2012), entre otras limitantes más que tal vez no se han atisbado.

## 1.2. *Web scraping*, o minería de red, para la recolección automatizada de datos

La naturaleza y el origen de los datos requeridos para el desarrollo de esta investigación implicarían un proceso de recolección sumamente lento y dificultoso: la sola actividad de revisar cada página electrónica de las plataformas dedicadas a la microfinanciación de proyectos, así como la selección, la clasificación y la extracción manual de la información necesaria (por medio del ya tradicional *copy & paste*) serían tareas que podrían tomar varias semanas. Para administrar de una manera más eficiente los tiempos para el acopio de los datos se optó por una alternativa de extracción automatizada de información de sitios web conocida como *web scraping* o “raspado de red”.

También denominado como minería de datos web (*web mining*), el *web scraping* es considerado tanto un tópico emergente como una actividad cada vez más demandada en el marco de la vertiginosa transformación digital que experimentamos hoy en día (László Hajba, 2018). Las aplicaciones principales del raspado de red dependen del tipo de información que se desea obtener, por lo que se pueden clasificar en: (a) minería del contenido de la web (*web content mining*), que comprende a su vez el minado de textos (*text mining*), el minado de hipervínculos (*hypertext mining*) y el minado de texto marcado (*markup mining*); (b) minado multimedia (*multimedia mining*); (c) minería de la estructura de la web (*web structure mining*) y (d) minería del uso de la web (*web usage mining*), que incluye el minado de redes sociales (*social web mining*) (García Herrero *et al.*, 2018).

De manera simplificada, el raspado de red se define como “la recolección automatizada de datos del internet”<sup>3</sup> (vanden Broucke y Baesens, 2018, p. 3); en detalle, el proceso en sí comprende la creación de un programa computacional que lleva a cabo, de manera automática y en mucho menor tiempo, la navegación, la recolección y la descarga de datos de numerosas páginas electrónicas, actividades que a una persona le exigirían lapsos mayores de dedicación. El atractivo que al día de hoy posee el raspado de red radica en

<sup>3</sup> “The automated gathering of data from the Internet”.

que el internet abarca una cantidad exorbitante de datos, estructurados y mayormente no estructurados (Mitchell, 2018), y cuyo uso va desde el análisis de indicadores estadísticos (por ejemplo, tasas de natalidad y mortandad de un país), la predicción de tendencias en redes sociales, modelos predictivos o análisis de opinión (*opinion mining*) para productos comerciales hasta el análisis de sentimientos (*sentiment analysis*) para servicios o campañas publicitarias y/o políticas (Russell y Klassen, 2019; vanden Broucke y Baesens, 2018). Además del raspado de red, que automatiza las tareas del internauta, se puede acceder a los múltiples datos de un sitio web a través de su interfaz de programación de aplicaciones (API), que consiste en rutinas programáticas para que terceros puedan acceder a la información de la página de manera estructurada; páginas como Twitter, Reddit, Facebook, LinkedIn,<sup>4</sup> entre otras, brindan un acceso restringido o libre a sus datos a través de sus respectivas API, ya sea de manera directa o mediante algún lenguaje de programación (Russell y Klassen, 2019), como Python (que es el empleado en este libro) o Java. La diferencia específica entre el uso de una API y el raspado de red consiste en que la primera supone un acceso (restringido, en varios casos) a la información de la página, brindado por esta misma, mientras que el raspado de red emula la interacción humana, mediante el navegador, con la información almacenada en el sitio web de interés (László Hajba, 2018; Mitchell, 2018). Además, el raspado de red es una alternativa viable en los casos en que la página no cuente con una API o ésta tenga algún costo o no brinde el acceso a la información que se desea obtener (vanden Broucke y Baesens, 2018).

A modo de corolario a este resumen sobre el raspado de red se puede profundizar en algunos de los usos prácticos que actualmente se le dan en diversos sectores de la actividad humana. Como se puede inferir de todo lo previamente expuesto, uno de los usos principales del raspado de red es el de la obtención de datos para su análisis posterior, esto es, como fuente de información para la ciencia de datos (László Hajba, 2018); así mismo, se emplea para alimentar bases de datos para tareas de procesamiento de lenguaje natural

---

<sup>4</sup> Desde el 2019 y a raíz de problemas de seguridad en la protección de datos de sus usuarios o por polémicas sobre el uso de los mismos como medio de lucro para otras empresas, Facebook, Instagram y LinkedIn han restringido los usos de sus API. Algunas páginas permiten el uso sólo académico o para fines de investigación, mientras que otras más incentivan incluso el uso comercial de sus datos mediante sus API.

tales como la clasificación de textos, como en Gutenberg Project<sup>5</sup> (Nelli, 2018), o para la traducción automática, como es el caso de Google Translate. En el sector bancario y de seguros, lo mismo que en el farmacéutico, la minería de red se emplea para analizar oportunidades de inversión, para predecir la morosidad de un cliente o bien para examinar la posibilidad de introducción de un nuevo fármaco en el mercado (García Herrero *et al.*, 2018).

En lo que respecta a la investigación en las humanidades y las ciencias sociales, el raspado de red se utiliza ya sea para la recolección de datos referentes a hábitos, consumo o prácticas específicas de la vida *online* (Li, Zhou y Cai, 2019), o bien para exploraciones sociodemográficas (Moro, Batista, Rita, Oliveira y Ribeiro, 2019), estudios psicológicos (Rafaeli, Ashtar y Altman, 2019) o para el análisis de la conducta política, como es el caso de los numerosos artículos relacionados con la respuesta del electorado a la actuación política de un ex presidente de los Estados Unidos (Martin, 2017). Por su parte, y luego de “minar” diversas imágenes de Tinder e Instagram con base en sus *likes*, Andrej Karpathy (2015) elabora un modelo predictivo de imágenes que serán más “atractivas” para los usuarios. Finalmente, y aparejado a los fenómenos del arte en red, del *net.art* y del arte virtual (Gronlund, 2017), se encuentra lo que se podría denominar como *data art*, que son creaciones artísticas basadas en la visualización de datos, como es el caso del proyecto *wefeelfine.org*, de Jonathan Harris y Sep Kamvar, el cual consiste en el mapeo y visualización de la frase inglesa *I feel* a través del minado de múltiples blogs (vanden Broucke y Baesens, 2018).

Según el objetivo y el tipo de información que se busca recopilar, así como la estructura del sitio web de interés, Python ofrece una amplia gama de librerías para el minado web. Una de las librerías más utilizadas —por su flexibilidad— es Beautiful Soup, que permite extraer el contenido de una página electrónica basada en el lenguaje de marcado HTML (HyperText Markup Language) y organizarlo en estructuras de datos legibles (László Hajba, 2018; Mitchell, 2018). La librería recibe su nombre del poema que aparece en *Alicia en el país de las maravillas* (Carroll, 2005) y alude al enmarañado conjunto de elementos que constituyen la estructura de una página web basada en HTML (ver figura 4).

<sup>5</sup> Gutenberg Project es una iniciativa abierta y gratuita que funciona como un repositorio o librería gratuita y que cuenta a la fecha con más de 60 000 títulos clasificados por autor, tema, idioma, etcétera.



más complejos (como la autocarga de contenido por deslizamiento) (vanden Broucke y Baesens, 2018).

Luego de examinar las páginas de crowdfunding de cobertura latinoamericana y en idioma español, tales como Ideame, Fondify y Mi cochinito, se seleccionó la primera para la obtención de datos que constituyen la muestra representativa de este proyecto, ya que casi la totalidad de los proyectos registrados en esta plataforma son de origen latinoamericano, además de que cada uno de éstos cuenta con métricas útiles para su análisis, como el monto de capital recaudado, el número de colaboradores, entre otras. Fondify y Mi cochinito fueron descartadas, ya que no mostraban datos suficientes para los objetivos de este estudio. En el caso particular de Kickstarter, a diferencia de Ideame, se observó que existía una mayor participación de creadores y creadoras mexicanas, en comparación con otros países latinoamericanos, motivo por el cual se decidió separar el análisis en los proyectos culturales y creativos de origen mexicano respecto del de los de América Latina en general. Cabe señalar que el conjunto de datos de Kickstarter no fue obtenido mediante un proceso de minería web, sino que se consiguió de la página de la compañía Web Robots, dedicada al raspado de red para empresas.

Para la recolección de datos de Ideame, se empleó Selenium para automatizar la navegación por categorías, así como para la carga de la página y para mostrar la totalidad de los proyectos según la categoría, lo mismo que para la extracción de sus datos y métricas. Es importante señalar que la recopilación del conjunto de datos requerido de Ideame se basó en los flujos de trabajo desarrollados por Mitchell (2018), y vanden Broucke y Baesens (2018); además, considerando la velocidad de navegación y la discriminación de datos no relevantes, este proceso se realizó en un lapso de aproximadamente cinco horas. Una acotación importante al respecto del proceso de minado de la página: la recomendación de diversos especialistas (László Hajba, 2018; Mitchell, 2018; vanden Broucke y Baesens, 2018) consiste en examinar primero el archivo robots.txt de la página que se desea minar, pues en éste se compendian los protocolos de exclusión o, en términos sencillos, indica si se puede acceder o no a la información de la página mediante herramientas de minería web. Para automatizar el análisis del archivo robots.txt se siguió el flujo de trabajo propuesto por László Hajba (2018), empleando el módulo urllib.robotparser.

Una vez seleccionados los datos necesarios para la investigación se empleó Pandas (The pandas development team, 2020), librería de Python enfocada en la manipulación y análisis de datos, para guardar la información recabada en un archivo de extensión .csv (*comma-separated values*) para su posterior procesamiento, análisis y visualización. Se obtuvieron siete conjuntos de datos correspondientes a los proyectos de las disciplinas de las artes escénicas, las artes visuales, cine y video, diseño, editorial, juegos y música, los cuales posteriormente se integraron en un solo archivo. Los datos recabados pertenecen a seis países latinoamericanos (ver tabla 1) e incluyen el título del proyecto, el nombre del autor/autora, el país correspondiente, una breve descripción del proyecto, el monto total recaudado (en pesos mexicanos), el porcentaje de cumplimiento del objetivo de recaudación, el total de visualizaciones y el de colaboradores (ver tabla 2).

Tabla 1. Relación de proyectos por país en el conjunto de datos de Ideame

<i>País</i>	<i>Artes escénicas</i>	<i>Artes visuales</i>	<i>Cine y video</i>	<i>Diseño</i>	<i>Editorial</i>	<i>Juegos</i>	<i>Música</i>
Argentina	170	146	413	49	270	17	254
Brasil	13	6	6	0	9	1	39
Chile	79	22	86	6	43	2	78
Colombia	3	4	10	2	7	0	1
México	2	8	18	17	7	1	0
Uruguay	6	3	8	1	2	0	15

Tabla 2. Vista de las variables del conjunto de datos de Ideame

	<i>Proyecto</i>	<i>Autor(a)</i>	<i>País</i>	<i>Descripción</i>	<i>Recaudación</i>	<i>Porcentaje</i>	<i>Vistas</i>	<i>Colaboradores</i>	<i>Categoría</i>
0	VIOLENTINA - UM JOGO MARGINAL DE CONTAR HISTÓRIAS	Eduardo Caetano	Brasil	Publicação do RPG Violentina, um jogo sobre Violência, Vícios, Volúpia... e nem um Vintém.	37172	459	6600	116	Editorial
1	CARD GOBLINS	Coisinha Verde Games	Brasil	Um incrível jogo de cartas onde cada jogador controla uma horda de caóticos e fracotes goblins!	44100	363	9300	225	Editorial

2	CHACABUCO, LA PELÍCULA	CHACABUCO	Argentina	Somos un grupo de estudiantes de Diseño de Imagen y Sonido (FADU – UBA). CHACABUCO es nuestra ópera prima y tesis final de la carrera.	8177	357	25300	851	Cine y video
3	DUNGEON MONSTERS: THE CARD GAME	Helio Greca	Brasil	Aqui não basta matar o monstro mais poderoso. Você deve matar todos eles - e antes dos seus oponentes!	45353	320	11900	204	Editorial
4	EL DISCO DE LA CHUCK NORRIS!!!	Alfonso Salas	Argentina	La Chuck Norris Band grabó su primer disco y sólo nos falta HACER las copias! Sé nuestro productor!!! Ofrecemos muchas recompensas!	59113	302	18100	100	Música
5	NO LO TESTEAMOS EXPANSIONES	Bureau de Juegos	Argentina	Dos nuevas expansiones del No Lo Testeamos Ni Un Poco: KINÉTICO, con cartas que te van a hacer mover y PICANTE, la expansión "hot".	19699	301	48400	395	Editorial

Como se comentó antes, los datos de Kickstarter no fueron minados directamente de su página por el autor de este estudio, sino que fueron obtenidos de un repositorio de la compañía Web Robots. En un principio se contó con 57 archivos .csv, los cuales contenían el título del proyecto, su descripción, el nombre del creador o creadora y su país de origen, así como su ubicación geográfica, la categoría a la que pertenece, la meta por recaudar y la recaudación final, el número de patrocinadores y el estatus del proyecto, entre otras categorías. Tras un examen preliminar de los diferentes ar-

chivos se observó que los países latinoamericanos con mayor presencia en Ideame tenían una participación mucho menor en Kickstarter, mientras que creadores y creadoras mexicanas cuentan con una mayor intervención en esta plataforma, motivo por el cual se acotó el análisis de los datos de Kickstarter a proyectos de origen mexicano. Si bien ambos conjuntos de datos poseen información específica que los diferencia, del análisis y comparación de ambos se puede obtener un panorama generalizado de la producción cultural latinoamericana.

Es necesario acotar que la muestra examinada en este estudio se obtuvo mediante un muestreo no probabilístico (Bernal, 2010), dado que se seleccionaron de manera deliberada los países de origen de los proyectos, descartando aquellos situados en los Estados Unidos, en el caso de Ideame, o los de cualquier otro país angloparlante, en lo que respecta a Kickstarter; los proyectos que conforman la muestra pertenecen a países englobados en la región geopolítica que recibe el nombre de Latinoamérica (García Zaballo e Iglesias Rodríguez, 2017). Es importante agregar que existe una serie de controversias teóricas respecto al uso y extensión del término Latinoamérica, o América Latina, entre las que se destacan la de los sesgos mediante los cuales se diferencia a una cultura como latina respecto de una angloamericana (Szurmuk y Mckee, 2009); los intereses que guían esta investigación conllevan a evitar este debate, asumiendo como válida la acepción más común de Latinoamérica, que comprende a las naciones que fueron colonizadas por países cuya lengua materna deriva del latín.

La muestra que se obtuvo de Kickstarter, por su parte, comprende únicamente proyectos culturales y creativos de procedencia mexicana; la categorización de la página Kickstarter es más amplia que la de Ideame, dado que distingue entre actividades tales como la danza y el teatro, que usualmente se clasifican como artes escénicas, lo mismo que separa la fotografía de las artes visuales. En la tabla 3 se puede observar la distribución de proyectos conforme a los estados o entidades federativas mexicanas, así como según las prácticas creativas a las que se corresponden. Es relevante indicar que en el conjunto de datos de Kickstarter las descripciones de algunos de los proyectos fueron elaboradas en inglés, probablemente para la captación de patrocinadores con mayor poder adquisitivo; en lo que respecta al conjunto de datos de Ideame, algunos proyectos fueron descritos en portugués,

por lo que en ninguno de los casos se cuenta con una muestra lingüísticamente uniforme.

Tabla 3. Relación de proyectos por estado en el conjunto de datos de Kickstarter

<i>Ubicación (por estado)</i>	<i>Cine y</i>										
	<i>Arte</i>	<i>Artesanía</i>	<i>video</i>	<i>Cómic</i>	<i>Danza</i>	<i>Diseño</i>	<i>Editorial</i>	<i>Fotografía</i>	<i>Juegos</i>	<i>Música</i>	<i>Teatro</i>
Aguascalientes	2	1	9	1	1	0	2	0	1	2	2
Baja California	56	14	147	24	6	15	51	13	23	62	24
Baja California Sur	1	0	0	0	0	0	0	1	0	0	1
Campeche	0	0	4	1	0	3	1	0	8	1	1
Chiapas	6	3	5	0	0	1	3	2	0	1	1
Chihuahua	6	2	3	3	4	1	2	0	3	7	0
Coahuila de Zaragoza	1	0	6	1	0	0	3	0	3	6	2
Colima	0	0	1	0	0	0	1	0	0	1	1
Distrito Federal	76	26	182	20	8	10	63	18	12	96	58
Durango	1	2	1	0	0	1	2	0	1	2	0
Guanajuato	7	3	24	0	0	3	6	7	4	8	2
Guerrero	1	0	1	0	0	0	0	0	0	1	0
Hidalgo	1	0	1	0	0	0	3	0	4	2	1
Jalisco	12	3	23	2	3	0	8	1	5	8	5
México	6	7	24	11	0	6	12	1	5	15	3
Michoacán de Ocampo	3	0	6	0	0	1	2	0	0	6	1
Morelos	2	0	3	0	0	0	5	0	1	3	3
Nayarit	1	0	4	0	0	0	0	0	0	1	0
Nuevo León	10	2	15	2	2	2	8	1	2	14	0
Oaxaca	1	1	4	0	0	0	2	1	0	1	2
Puebla	8	6	21	2	2	1	5	1	7	6	3
Querétaro de Arteaga	4	3	24	0	0	10	5	2	3	6	2
Quintana Roo	8	2	7	1	0	0	3	2	0	2	0
San Luis Potosí	1	2	5	4	2	1	5	0	2	6	1
Sinaloa	0	8	2	2	3	0	2	0	0	1	0
Sonora	3	3	4	0	0	0	7	1	0	3	0
Tabasco	2	0	1	2	0	0	2	1	0	1	0
Tamaulipas	4	1	4	0	0	0	1	1	0	5	0
Veracruz-Llave	1	6	19	1	0	0	7	4	5	12	2
Yucatán	2	8	6	2	1	1	7	3	3	7	3
Zacatecas	0	0	8	1	0	1	4	1	1	0	0

Independientemente de si los datos de investigación se obtuvieron mediante técnicas “tradicionales” de recolección, como encuestas, entrevistas, sondeos, etc. (Hernández Sampieri, Fernández Collado y Baptista Lucio,

2014), o mediante técnicas asistidas por alguna tecnología, tales como formularios electrónicos, bases de datos digitales o, como es el caso de esta investigación, a través del minado de la web, los datos recabados requieren ser procesados a través de diferentes procedimientos que garanticen su legibilidad (Bernal, 2010). En el análisis de datos el procesamiento de los mismos es también conocido como fase o etapa de preparación (García Herrero *et al.*, 2018; McKinney, 2018; Nelli, 2018) y comprende las actividades de limpieza, filtrado, transformación, estandarización, eliminación, combinación y organización (Embarak, 2018; Henley y Wolf, 2018). En el caso particular que aquí se aborda, ambos conjuntos de datos fueron procesados para eliminar inconsistencias y/o valores repetidos o ambiguos, adecuándolos lo más posible para su lectura y análisis.

Para el conjunto de datos obtenido de Ideame, se llevó a cabo primeramente una comprobación de valores nulos,<sup>7</sup> seguida de una comprobación relativa a los valores duplicados, los cuales, una vez localizados y descartados, redujeron el tamaño del conjunto general de proyectos de 3 757 observaciones a 1 824, y dado que sólo se identificó la existencia de dos valores nulos, se pudo verificar que una cantidad considerable de datos era repetida; este proceso de filtrado y eliminación corresponde a una limpieza preliminar de los datos de investigación (McKinney, 2018) que es complementada con una operación de transformación para convertir ciertos valores a otro tipo que facilite o simplifique su análisis: en el conjunto de datos de Ideame, las métricas de los proyectos (recaudación, porcentaje, vistas y colaboradores) que aparecían como cadenas de caracteres alfabéticos se transformaron a números enteros, posibilitando un examen más eficiente (Nelli, 2018). Tras la limpieza, eliminación y transformación, los datos fueron integrados en un nuevo conjunto organizado en orden descendente según el porcentaje de cumplimiento.

Como se comentó anteriormente respecto a los datos de Kickstarter, se realizó primero la integración de los diferentes conjuntos en uno solo, e inmediatamente se procedió a examinar la participación de los creadores según su país de origen, comprobándose la primacía de los pertenecientes a países angloparlantes; México fue el único país de América Latina presente

---

<sup>7</sup> Los cuales pueden aparecer como NaN (Not-a-Number) o Null.

en el conjunto general. Tras esta primera integración, se creó un segundo conjunto de datos específico para los proyectos de origen mexicano, reduciendo y filtrando los datos a un total de 2 917 filas y 41 columnas, para después buscar los datos nulos y los duplicados, que fueron eliminados junto con las columnas no relevantes para la investigación. Una vez realizadas las operaciones de reducción y discriminación, se conservaron 1 841 filas y nueve columnas correspondientes al nombre del creador/creadora, su ubicación geográfica, el título y una breve descripción del proyecto, su categoría o disciplina, el objetivo fijado y la cantidad recaudada final, el número de patrocinadores y el estatus del proyecto (fallido, exitoso, en curso, cancelado) (ver tabla 4).

Tabla 4. Vista de las variables del conjunto de datos de Kicstarter

<i>Conteo de patrocinadores</i>	<i>Descripción</i>	<i>Objetivo</i>	<i>Nombre del proyecto</i>	<i>Estatus</i>	<i>Ubicación (por estado)</i>	<i>Recaudado</i>	<i>Creador(a)</i>	<i>Categoría</i>
0 2919	A Heartfelt Metroidvania Experience	380000	9 Years of Shadows	Exitoso	Jalisco	1988905	HALBERD STUDIOS	Juegos
1 1990	The Kinetic Sculpture That Draws Infinite Patterns On Sand.	100000	SANDSARA	Exitoso	Nuevo León	18574946	Ed Cano	Diseño
2 1806	A limited edition deck of playing cards that celebrates life and death in the Mexican culture	280000	El Reino de Los Muertos - Expert Edition / Mexico Creates	Exitoso	Distrito Federal	1322488	Sextante studio	Juegos
3 1339	Come and check out our innovative, simple and practical luggage organizer. Hasta la vista, messy clothes!	150000	Hang and Roll "Organize your backpack like never before"	Exitoso	Baja California	1039589	Hang & Roll	Artesanía

4	1132	2020 not only takes us back into the studio for production of our 3rd album, but we embark on our first North American Tour March/April	570000	The Warning - New Album and 23 date NorthAmerican Tour	Exitoso	Nuevo Leon	3337199	The Warning	Música
5	1075	Bring your Tabletop sessions to life. Immersive auditive experiences for Dungeon Masters and Players at your fingertips.	369000	RPG Sounds	Exitoso	Baja California	1338164	Cast n Play	Juegos

Además del procesamiento previamente descrito, en las columnas del nombre del creador, de la ubicación geográfica y de la categoría artística o creativa del proyecto se llevaron a cabo tareas adicionales de limpieza para eliminar texto no relevante o caracteres que dificultaran la revisión de los datos, y posteriormente se reorganizó la información en orden descendente en función del número de patrocinadores; por último, se realizó una comprobación final de valores nulos y/o duplicados para evitar inconsistencias durante el análisis. Es importante reiterar que, a diferencia de los datos recabados de la página Ideame, los de Kickstarter cuentan con categorías más específicas de disciplinas artísticas y creativas, que comprenden el arte (que incluye la pintura y la escultura), las artesanías, el cine y el video, el cómic, la danza, el diseño, el trabajo editorial, la fotografía, los juegos, la música y el teatro.

Conviene subrayar que la categorización de cada página puede considerarse, en cierta medida, como un indicativo de la organización material del campo cultural, es decir, de los medios y procedimientos de producción simbólica, así como de sus relaciones sociales (García Canclini, 2014). En otras palabras, una clasificación como la de Kickstarter apuntaría a un posicionamiento del proyecto en función del mercado cultural, mientras que una categorización como la de Ideame (que integra, por ejemplo, en artes

visuales a la pintura, la fotografía, la novela gráfica y la escultura, o, en artes escénicas a la danza, el *performance* y el teatro) tendería hacia un posicionamiento centrado en el producto cultural mismo (Colbert y Cuadrado, 2015).

### 1.3. Análisis de datos del micromecenazgo cultural y creativo

Luego del preprocesamiento preliminar y la consecuente reducción de datos antes comentada, se lleva a cabo un análisis exploratorio (EDA) que consiste en un proceso de indagación inicial en torno a los datos recabados con el objetivo de ratificar o rechazar, según sea el caso, los supuestos sobre los que se basa la investigación en curso (Henley y Wolf, 2018). El análisis exploratorio, como aproximación primaria a los datos, se apoya en técnicas estadísticas básicas, así como en la graficación de la información, para descubrir patrones o tendencias o descartar datos innecesarios (Patil, 2018). En este estudio, el análisis exploratorio tiene como propósito confirmar si del examen de los datos de los proyectos recolectados de las páginas de Ideame y Kickstarter, se puede obtener información relevante para el desarrollo, posicionamiento y promoción de campañas de apoyo y de financiamiento para proyectos artísticos, culturales y creativos en general.

Una de las primeras actividades del EDA consiste en identificar los tipos de variables presentes en los conjuntos de datos examinados. En lo que respecta a los datos obtenidos de Ideame, se cuenta con cuatro variables cuantitativas o numéricas discretas, esto es, cuyos valores se conforman por números enteros (Patil, 2018), como lo son la recaudación, el porcentaje de cumplimiento, las vistas y el número de colaboradores que tuvo cada proyecto; las variables cualitativas o categóricas de este primer conjunto de datos corresponden al nombre y una descripción breve del proyecto, su autor, el país de origen y la categoría en la que se incluye, las tres primeras son de carácter nominal, lo que implica que sus valores no están organizados en torno a un orden explícito o que poseen una alta cardinalidad (Stepanek, 2020), y las dos últimas de carácter ordinal, puesto que se puede inferir una clasificación intrínseca en los valores representados (Henley y Wolf, 2018).

Por su parte, el conjunto de datos recabado de Kickstarter cuenta con tres variables numéricas, también discretas, que son el conteo de patrocinadores, el objetivo propuesto y la recaudación final de cada proyecto. De las variables categóricas, las referentes al nombre, la descripción y el creador del proyecto, al tener una gran cantidad de valores diferentes, cuentan con una alta cardinalidad (McKinney, 2018), mientras que las variables del estatus, la categoría del proyecto y su ubicación por estado de la República, son de carácter ordinal. Una vez identificados los tipos de variables, el siguiente paso es examinar las métricas básicas de los conjuntos de datos explorados, que incluyen las tendencias centrales (media, mediana y moda) y los percentiles, medidas estadísticas que resumen los valores relativos a las variables cuantitativas en la muestra de análisis (Foxwell, 2020). En las tablas 5 y 6 se presentan las métricas básicas de los datos de Ideame y de Kickstarter respectivamente.

Tabla 5. Métricas generales del conjunto de datos de Ideame

	<i>Recaudación</i>	<i>Porcentaje</i>	<i>Vistas</i>	<i>Colaboradores</i>
count	1826	1826	1826	1826
mean	21047.56407	71.51971522	9125.679628	84.52902519
std	54474.4978	46.53044836	12285.06923	206.8602559
min	16	10	504	1
25%	3093	29	3400	17
50%	7516	73	6200	37
75%	18720.75	104	10900	74
max	1565884	459	299600	2693

Tabla 6. Métricas generales del conjunto de datos de Kickstarter

	<i>Conteo de patrocinadores</i>	<i>Objetivo</i>	<i>Recaudado</i>
count	1854	1854	1854
mean	35.89536138	179822.4288	47935.26214
std	119.3471396	2368868.549	449311.0446
min	0	10	0
25%	2	20000	100
50%	11	45000	7396.5
75%	38.75	90000	38359
max	2919	100000000	18574946

Una de las principales ventajas de la observación de las métricas generales consiste en que permite establecer comparaciones valiosas referentes

a los valores de las variables numéricas de los conjuntos de datos examinados; en este caso, se puede comprobar que los proyectos culturales y creativos en Ideame alcanzan una recaudación máxima de más de un millón de pesos, con un porcentaje de éxito de más de 400%, con un número de vistas de casi 300 000, con más de 2 000 colaboradores; no obstante, en el caso de Kickstarter, el máximo recaudado es de casi 20 millones de pesos, con casi 3 000 patrocinadores. De esta sucinta comparación se puede inferir que esta página posee una mayor cobertura y popularidad que otras centradas específicamente en América Latina y México, como lo son Ideame y Donadora, por lo que las posibilidades de éxito de las campañas desarrolladas en Kickstarter son, por lo visto, mayores.

Después de identificar las variables de los conjuntos de datos analizados, así como de explorar sus métricas descriptivas, el siguiente paso del EDA es realizar un análisis univariado, que consiste en examinar cada una de las variables por separado, ya sea a través de la presentación tabular de los datos o bien mediante diferentes gráficos (Embarak, 2018). Dado que en los conjuntos de datos que conforman la muestra de este estudio predominan variables categóricas de alta cardinalidad, se opta por un análisis multivariado, que consiste en analizar las correlaciones entre dos o más variables de manera simultánea (McKinney, 2018); este tipo de análisis, que es además el siguiente paso del análisis exploratorio, es de utilidad cuando las variables del conjunto de datos examinado no poseen un orden claro o una clasificación apriorística (Patil, 2018; Stepanek, 2020).

En lo que respecta al conjunto de datos de Ideame, en las figuras 5, 6, 7 y 8 se visualizan las relaciones entre las variables de recaudación, porcentaje, vistas y colaboradores en función del país y de la categoría de los proyectos. Según los gráficos presentados, se observa que los proyectos de mayor recaudación son los de diseño, seguidos por los editoriales, de cine y video y de música; los de menor recaudación corresponden a los proyectos de artes visuales y escénicas, y de juegos. En lo que respecta al porcentaje de financiamiento, se percibe un patrón parecido, sólo que en este caso son los proyectos editoriales los que alcanzan el valor más alto, mientras que las demás categorías poseen valores semejantes, exceptuando los proyectos de juegos, que en todos los países no superan ni siquiera 50% de financiamiento. La gráfica relativa a las vistas permite observar que los proyectos de

diseño son los que cuentan con una mayor cantidad de visualizaciones, al lado de los de cine y video y los editoriales, mientras que los de menor visualización corresponden a los de artes escénicas.

Figura 5. Relación de recaudación por categoría según país (conjunto de datos de Ideame)

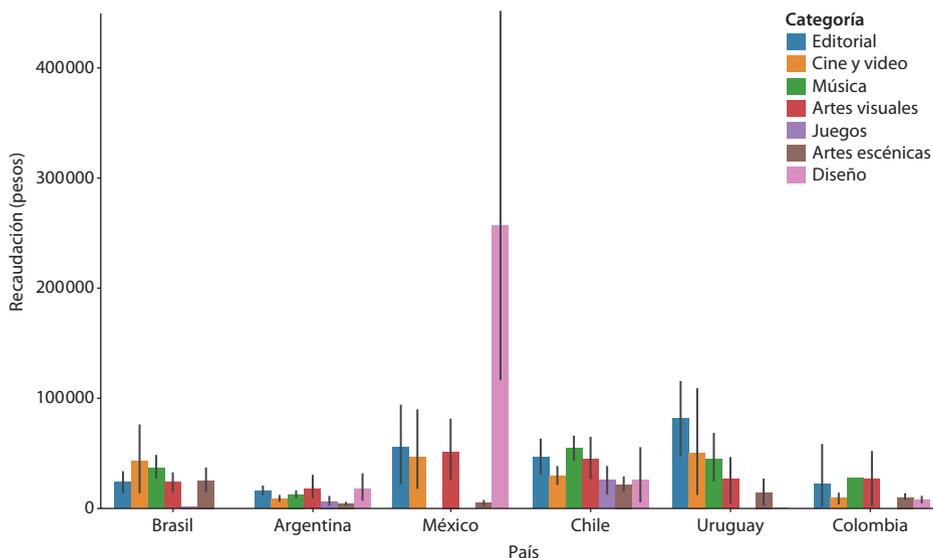


Figura 6. Relación de porcentaje de financiamiento por categoría según país (conjunto de datos de Ideame)

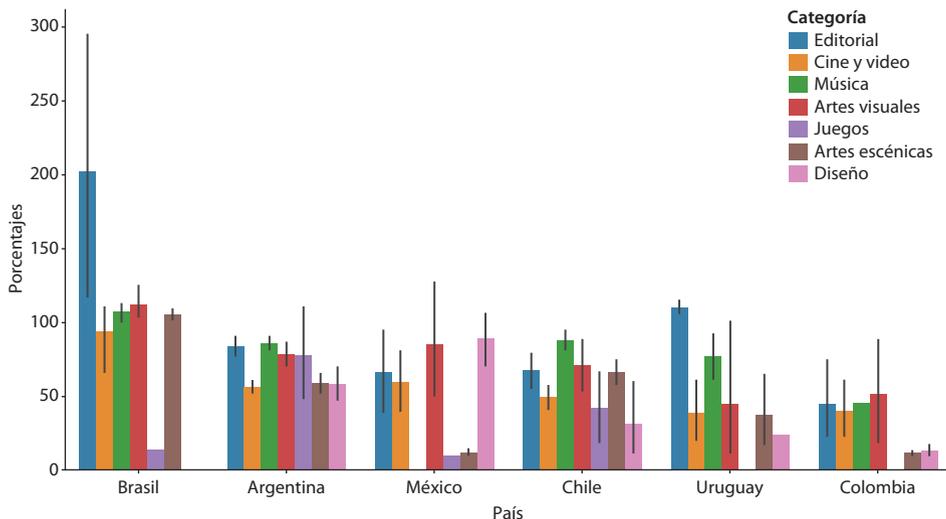


Figura 7. Relación de vistas por categoría según país (conjunto de datos de Ideame)

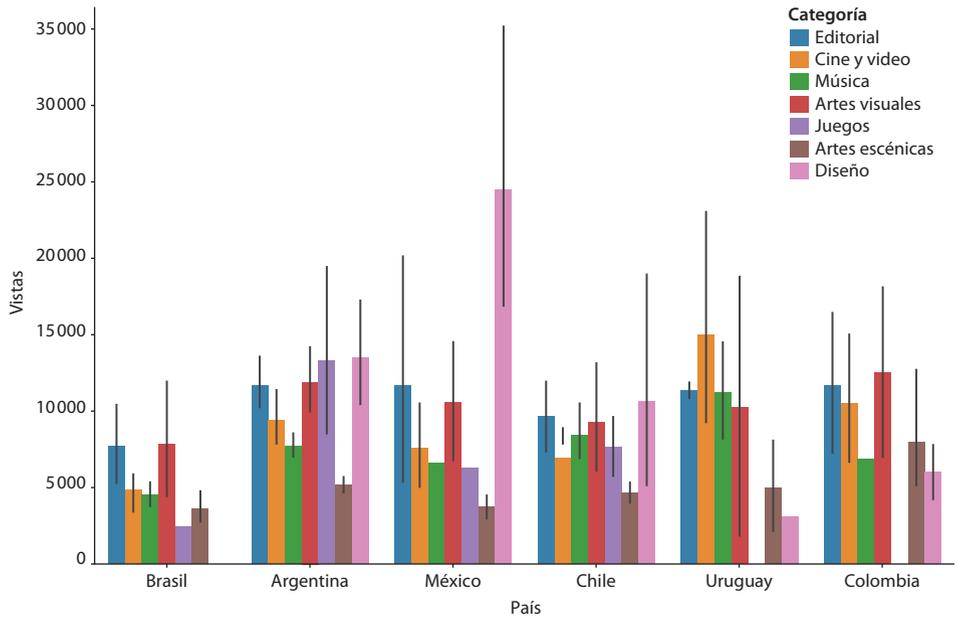
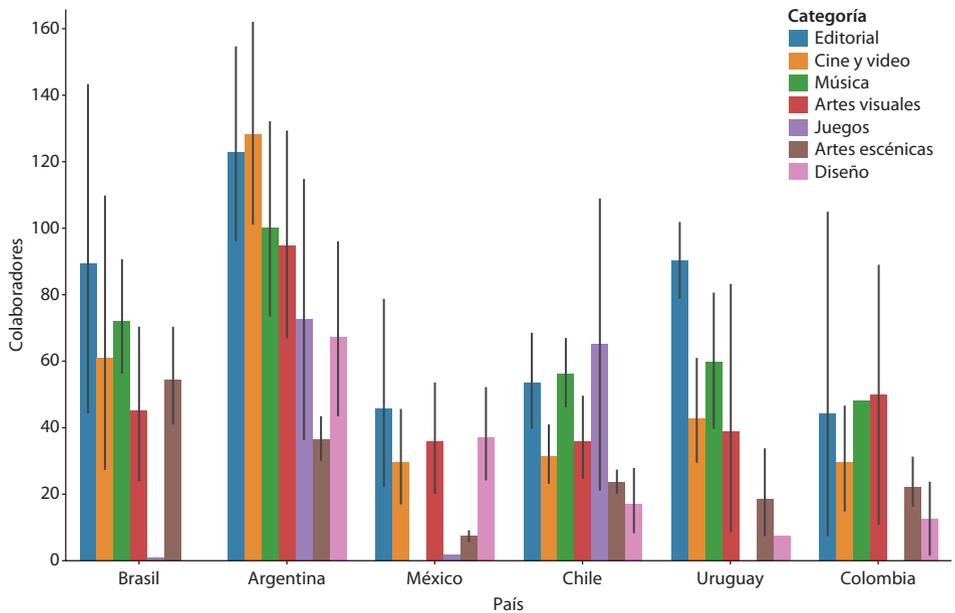


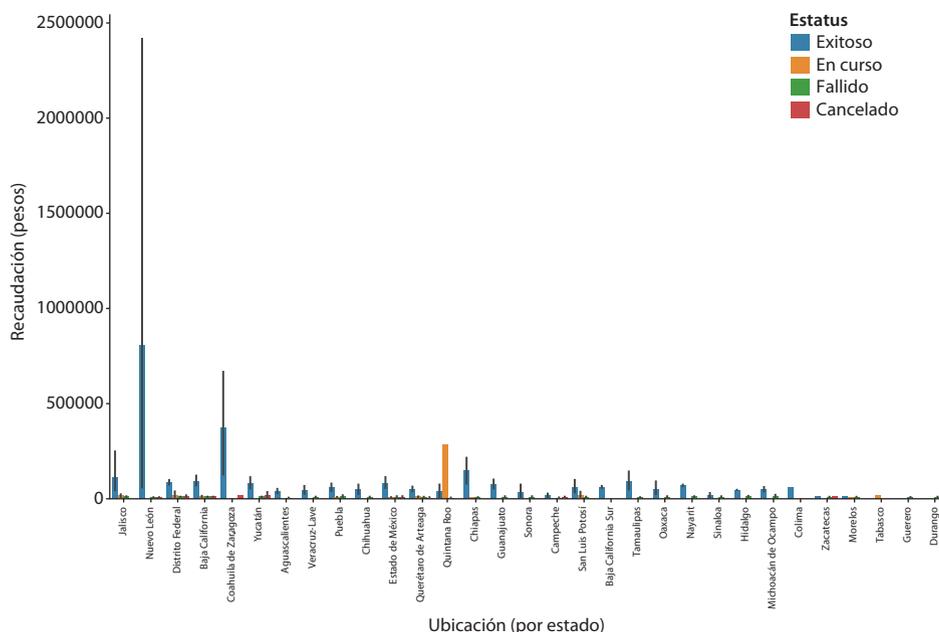
Figura 8. Relación de colaboradores por categoría según país (conjunto de datos de Ideame)



La última gráfica, en la que se visualiza la cantidad de colaboradores por categoría según el país, permite observar una mayor participación y respaldo a proyectos de cine y video, de música y editoriales, todos de origen argentino. A propósito de estas visualizaciones, se identifica una tendencia a apoyar principalmente proyectos que no requieren necesariamente de un espacio físico para su montaje o representación, como es el caso del cine y el video, y de los proyectos editoriales y musicales, dejando a las artes visuales y a las escénicas, desafortunadamente, con un muy bajo índice de respaldo. De igual manera, países como Argentina y Brasil parecieran tener una mayor intervención en el desarrollo y promoción de proyectos culturales y creativos a través de Ideame, lo que implica que, o bien existe un mayor número de productores de origen argentino y brasileño en esta plataforma o, en su defecto, este sitio web ha tenido un mayor éxito con usuarios de estos países.

Por su parte, en las figuras 9, 10, 11 y 12 se visualizan las correlaciones entre variables del conjunto de datos de Kickstarter correspondientes a la

Figura 9. Relación de recaudación por estatus según estado de la República (conjunto de datos de Kickstarter)



recaudación, el objetivo y el número de patrocinadores por estatus según el estado de la República, así como por categoría. Uno de los hallazgos observados a través de estos gráficos es, primero, que el máximo de recaudación en una campaña, como se mencionó anteriormente, llega a un tope de casi 20 millones de pesos mexicanos, lo que implica que se trata de una excepción o una anomalía, dado que el promedio en lo que respecta a este rubro es de menos de los \$100 000 pesos. Sobre el objetivo de recaudación se percibe un alto índice de campañas fallidas y canceladas con metas superiores a los \$100 000; en lo que respecta al número de patrocinadores, se observa que, como se podía esperar, las campañas exitosas cuentan con una cantidad superior a los 50 patrocinadores.

Figura 10. Relación de objetivo de recaudación por estatus según estado de la República (conjunto de datos de Kickstarter)

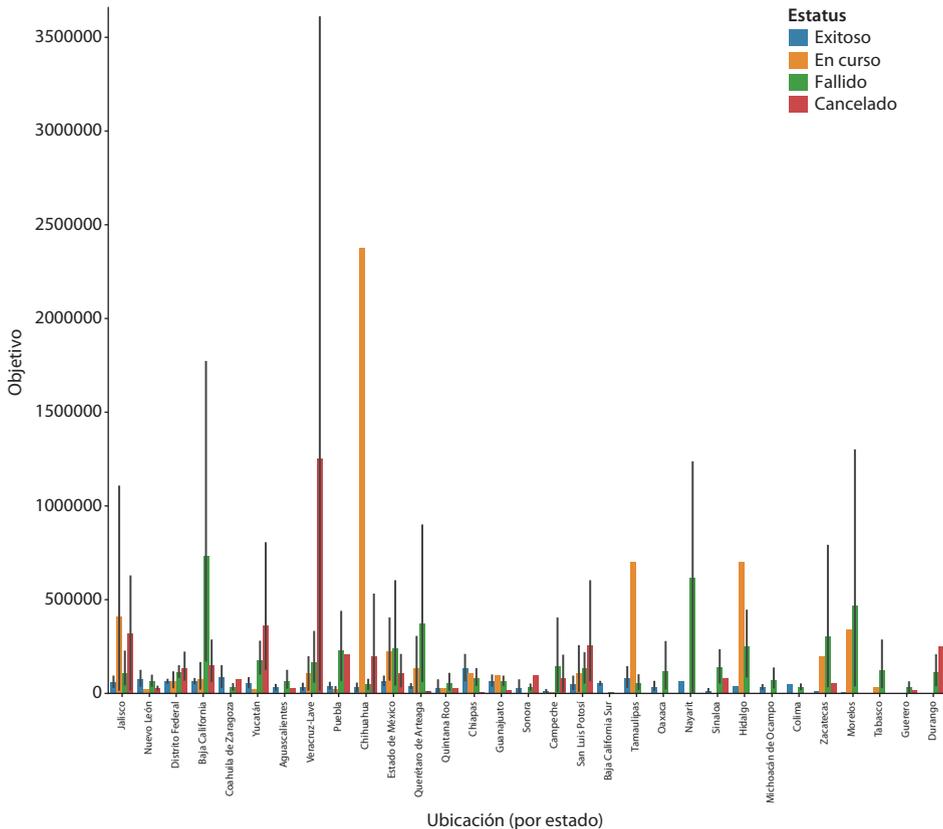


Figura 11. Relación de patrocinadores por estatus según estado de la República (conjunto de datos de Kickstarter)

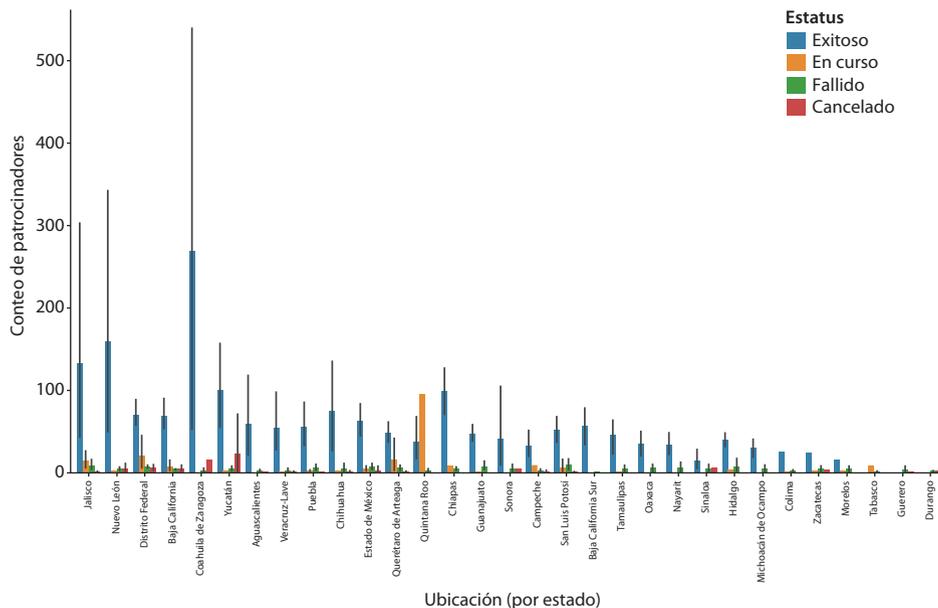
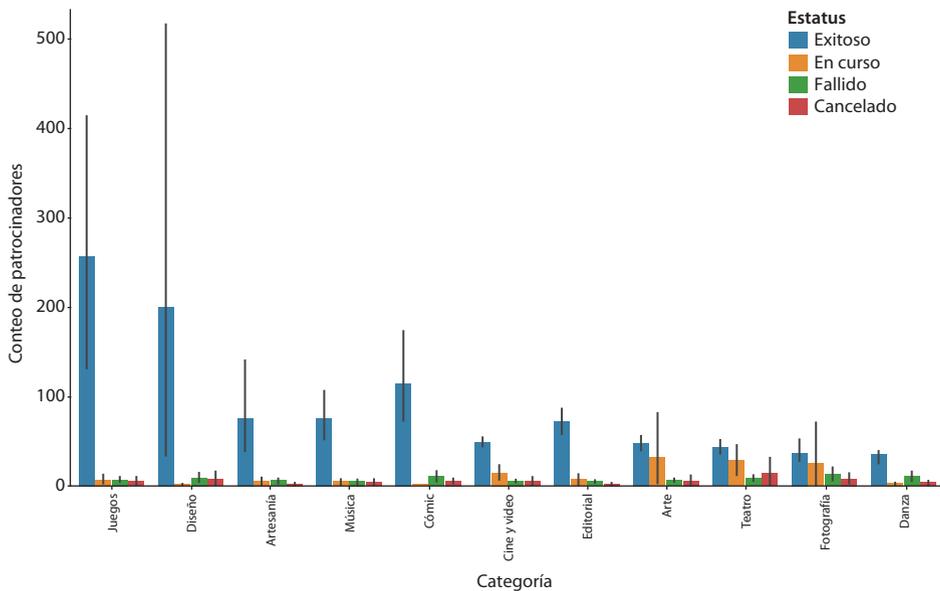


Figura 12. Relación de patrocinadores por estatus según categoría (conjunto de datos de Kickstarter)



En lo relativo a las categorías de proyectos apoyados es importante señalar que los que cuentan con mayor respaldo son los clasificados bajo las etiquetas de juegos y de diseño, seguidos por cómics, editorial, cine y video, música y artesanía, mientras que los que han tenido un menor patrocinio son los incluidos en las categorías de arte, teatro, fotografía y danza. Si se comparan los gráficos de las campañas realizadas en Kickstarter con los proyectos promocionados en Ideame se evidencian algunos datos de gran relevancia.

Lo primero que se debe destacar es que en ambos sitios los proyectos que han obtenido un mayor financiamiento son aquellos que pudieran considerarse como formas de creación cultural o prácticas creativas “populares” (Spracklen, 2015), como el cine, el video y el cómic, o funcionales, como el diseño; el problema que subyace a este hallazgo es el del rezago de las formas culturales consagradas, es decir, el poco apoyo que reciben proyectos o campañas centradas en las artes visuales y las escénicas. Un segundo aspecto que conviene mencionar es que en ambos conjuntos de datos se constata una correlación entre las variables referentes a los colaboradores o patrocinadores, la recaudación y el objetivo de financiamiento/estatus del proyecto; cabe preguntarse si las categorías de los proyectos también se encuentran relacionadas con las variables mencionadas. Para responder a esta interrogante se procede a continuación al siguiente paso del EDA, que consiste en el análisis de las correlaciones (Henley y Wolf, 2018).

El análisis correlacional permite explorar los nexos cuantitativos entre dos (o más) variables, lo que posibilita examinar la incidencia de la primera sobre la segunda (McKinney, 2018), es decir, cómo los cambios o variaciones en una variable afectan a otra. El análisis de correlaciones es uno de los pasos esenciales para construir un modelo predictivo, pues, además de auxiliar en la comprensión de los vínculos entre variables, es también de gran utilidad para identificar las causas de un fenómeno, así como para inferir sus efectos (Nelli, 2018). Para este estudio se emplean tanto el coeficiente de correlación de Pearson (*Pearson correlation coefficient*), que calcula las correlaciones entre variables cuantitativas o numéricas de manera lineal (Embarak, 2018; Stepanek, 2020), como el coeficiente de correlación Phik, que estima las correlaciones entre variables cuantitativas (discretas o continuas) y cualitativas o categóricas de tipo ordinal (Baak, Koopman, Snoek y Klous, 2019; Zychlinski, 2019).

Para el conjunto de datos de Ideame, el coeficiente de correlación Pearson permite observar (ver figura 13) una alta correlación entre la recaudación de fondos y las vistas de los proyectos, pero, de manera inusual, existe una correlación media con respecto a los colaboradores, lo que implica que la cantidad de colaboradores de un proyecto no garantiza la obtención de recursos; por otra parte, se percibe una alta correlación entre el número de colaboradores y las vistas, y, en menor medida, con el porcentaje de financiación, dado lo cual se puede afirmar que el visionado de un proyecto atractivo es un factor que determina la captación de colaboradores y, en una menor medida, el porcentaje con el que éstos apoyarían un proyecto. Por su parte, la matriz de correlación Phik (ver figura 14), además de corroborar los nexos previamente comentados, evidencia una muy baja correlación de las variables numéricas respecto al país de origen de los proyectos promocionados y una correlación de media a baja en lo que respecta a la categoría en la que se suscribe el proyecto, en relación con su número de vistas, su recaudación y el porcentaje de financiamiento.

Figura 13. Matriz de correlación Pearson (conjunto de datos Ideame)

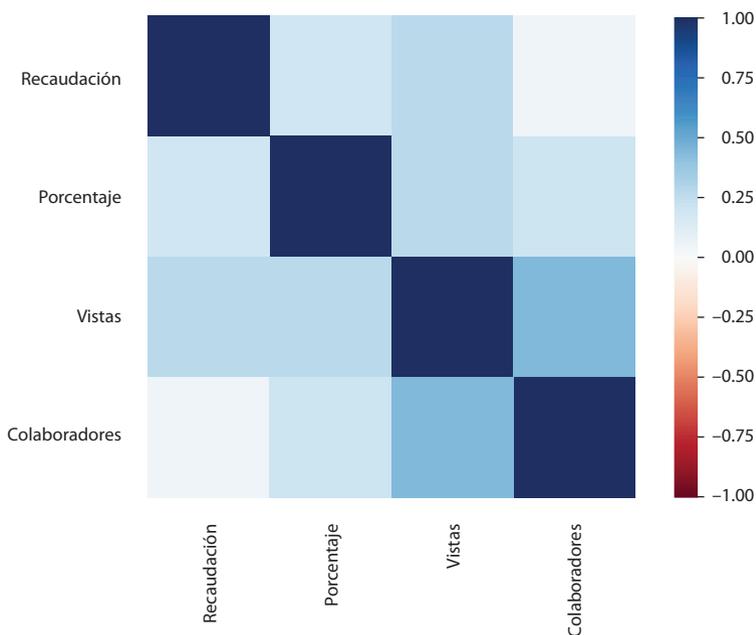
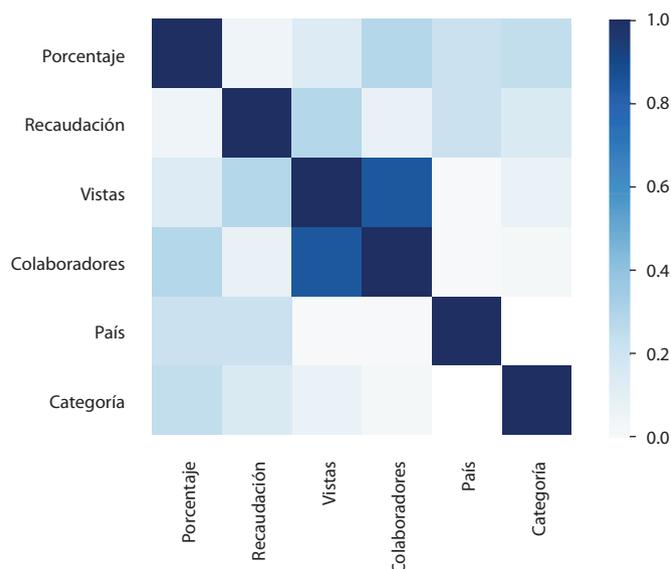


Figura 14. Matriz de correlación Phik (conjunto de datos Ideame)



Es de interés señalar que existe una correlación media entre la categoría de los proyectos y el país de procedencia, de lo que se puede concluir que existe una inclinación por parte de los usuarios del sitio hacia ciertas formas de producción cultural o prácticas creativas según el espacio geopolítico al que pertenezcan, confirmando en cierta medida la premisa anteriormente expuesta sobre una mayor participación de creadores y usuarios argentinos y brasileños en la página de Ideame. Por otra parte, y como se indicó antes, aunque existe una correlación de media a baja en cuanto a las variables de categoría, porcentaje y vistas, hay una muy baja correlación con respecto al número de colaboradores, lo que implica que la categoría a la que pertenece el proyecto no es un factor decisivo para la atracción de colaboradores.

En lo que respecta al conjunto de datos conformado por las campañas de Kickstarter, se observa una correlación de media a alta entre el monto recaudado por proyecto y la cantidad de patrocinadores, así como una correlación media entre estas variables y la del objetivo de financiación (ver figura 15); si bien los nexos entre estos parámetros podrían considerarse como evidentes, son también indicativos de que el cumplimiento de la meta de recaudación no depende principalmente del número de patrocinadores, sino que existen otros factores que seguramente inciden en esta variable. Al

posibilitar el examen de las relaciones entre variables categóricas y numéricas, la matriz de correlación Phik (ver figura 16) hace patente una correlación de baja a media entre la categoría de las campañas, el conteo de patrocinadores, su ubicación por entidad federativa, así como su estatus —si el proyecto es exitoso, está en curso, ha fallado o ha sido cancelado— y el monto final recaudado.

Figura 15. Matriz de correlación Pearson (conjunto de datos Kickstarter)

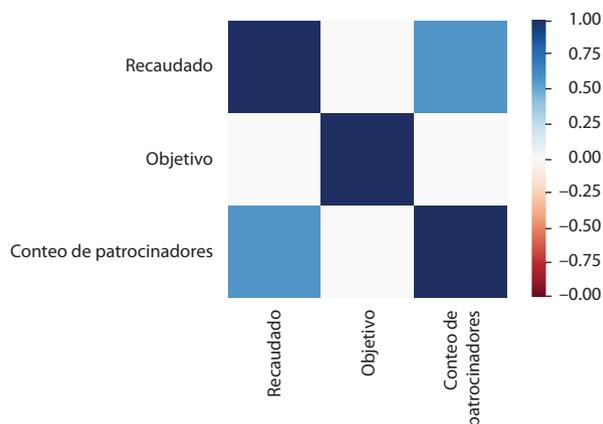
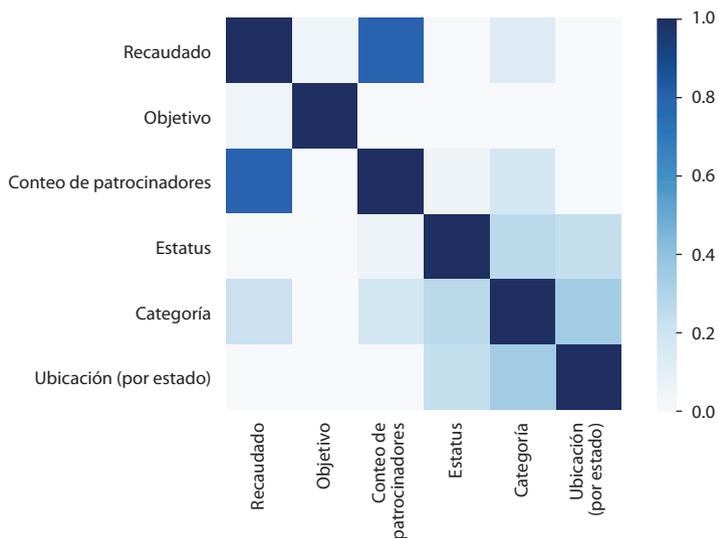


Figura 16. Matriz de correlación Phik (conjunto de datos Kickstarter)



De lo anterior se infiere que en algunos estados de la República existe una tendencia más acentuada a respaldar iniciativas culturales y creativas independientes a través de este sitio web, además de que hay una preferencia por ciertos tipos particulares de proyectos, siendo los juegos, el diseño y el cómic las categorías con mayores índices de apoyo y de éxito y, en concordancia, Jalisco, Nuevo León, Baja California, Coahuila, Puebla, Querétaro y la Ciudad de México son los estados que cuentan con un mayor registro de patrocinadores activos y de campañas concluidas exitosamente. Con el examen de las correlaciones que se ha llevado a cabo previamente se puede dar por terminado el análisis exploratorio, o EDA, cuyo propósito es el de brindar la información necesaria para modelar un sistema de recomendación con el que se pueda “predecir” el apoyo que los usuarios brindarían a un proyecto cultural y creativo.

Un sistema de recomendación es una de las aplicaciones más comunes de la ciencia de datos y consiste en una serie de algoritmos por medio de los cuales se realiza un filtrado de productos o servicios (libros, películas, canciones, noticias, vacantes laborales, etc.) basado en indicadores determinados para devolver un conjunto de opciones que se ajustan a las preferencias del usuario (Albon, 2018). Las sugerencias de visionado de series y películas en Netflix, las listas de reproducción en YouTube, el historial de productos vistos en Amazon o las recomendaciones de hospedaje en AirBnB son algunos de los ejemplos más populares de sistemas de recomendación en la actualidad; la premisa general que subyace a esta aplicación es que las preferencias de los usuarios son de utilidad para pronosticar sus elecciones futuras o las de otros posibles beneficiarios (Lope.ai, 2020).

En la industria se reconocen generalmente dos tipos de sistemas de recomendación: el de filtrado basado en contenido y el de filtrado colaborativo (Malik, s. f). El primero funciona estimando la similitud entre objetos según sus atributos, por ejemplo, si un libro o una película pertenecen a un género en particular, si una canción tiene una duración semejante a otra, etc. (Hug, 2020); el sistema de filtrado colaborativo, por otra parte, se basa en las interacciones de los usuarios con el objeto o evento: en función de las preferencias de un número de usuarios —expresadas mediante likes, retweets u otras formas de interacción digital— se extrae la probabilidad

con la que otros usuarios optarían por elegir el mismo producto u otros semejantes (Ajitsaria, s. f.).

Esta investigación culmina precisamente con el diseño de una aplicación web<sup>8</sup> que a su vez contiene dos sistemas de recomendación modelados a partir de los conjuntos de datos de Ideame y de Kickstarter; es necesario indicar, primero, que pese a que se encuentran en un estado todavía rudimentario, ambos sistemas de recomendación son escalables, es decir, se pueden adaptar con facilidad a las necesidades cambiantes de los usuarios, además de que su desarrollo admite grados mayores de cobertura, extensión y complejidad (Mathur, 2019); segundo, se elaboraron, por una parte, un sistema de recomendación simple, basado en la correlación de las métricas generales de cada conjunto de datos, y, por otra parte, un segundo sistema, un tanto más complejo, basado en las descripciones de los proyectos de Ideame y de Kickstarter, respectivamente. Ambos sistemas se clasifican como de filtrado basado en contenido, pues operan en función de las propiedades mismas de los objetos, antes que en las interacciones específicas de los usuarios con respecto a éstos (Albon, 2018).

Otro aspecto que es pertinente señalar es que para el desarrollo del segundo sistema de recomendación se efectuaron tareas de procesamiento de lenguaje natural en las descripciones de los proyectos en ambos conjuntos de datos: se eliminaron palabras vacías, o *stopwords* (como artículos, pronombres u otras partículas gramaticales con poco o nulo valor semántico), se eliminaron símbolos tales como paréntesis, corchetes y emoticonos (Dwivedi, 2020), y se extrajeron tres palabras clave por cada descripción para facilitar la estimación de similitudes en los contenidos textuales (Radim y Sojka, 2010). Así pues, a los conjuntos originales se añadieron dos variables, correspondientes a la descripción limpia y las palabras clave de cada proyecto. En el siguiente capítulo se profundizará en torno a las técnicas de NLP).

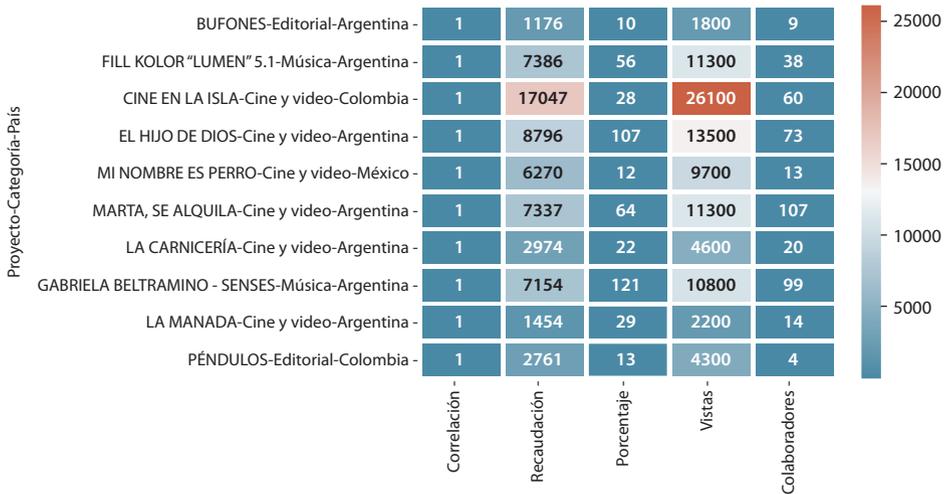
El funcionamiento del primer sistema de recomendación, como se indicó de manera sucinta anteriormente, se basa en la correlación comparada

---

<sup>8</sup> Se invita a lectoras y lectores interesados a revisar la aplicación web en el enlace que a continuación se comparte, para probar su funcionamiento, realizar sus propias comprobaciones, según sus intereses, y brindar retroalimentación para su mejora: <https://share.streamlit.io/jeoa-1981/apprecomendation/main/AppRecommendation.py>

de las métricas de cada proyecto; en otras palabras, se selecciona un proyecto en particular y el sistema “devuelve” nueve opciones semejantes en función de la recaudación, el objetivo de financiamiento, las vistas, los colaboradores o patrocinadores, según se trate del conjunto de datos de Ideame o de Kickstarter. En el siguiente ejemplo se realiza una comprobación del sistema con un proyecto del conjunto de datos de Ideame elegido al azar (ver figura 17); como se puede observar en el gráfico, los proyectos con los indicadores resaltados en gamas cromáticas rojizas —específicamente el que concierne a las vistas— son, teóricamente, los que tienen una mayor probabilidad de ser respaldados por los usuarios.

Figura 17. Resultados de recomendación basada en métricas (conjunto de datos Ideame)



En lo que respecta al segundo sistema de recomendación —que está basado en el contenido textual de ambos conjuntos de datos— se puede verificar su funcionamiento seleccionando el mismo proyecto del conjunto de datos de Ideame; la diferencia sustancial con respecto al primer sistema de recomendación radica en que, en este caso, no son únicamente las métricas las que influyen en las sugerencias devueltas, sino también la similitud entre descripciones, así como la frecuencia de las palabras clave presentes en cada proyecto, como se puede observar en la nube de palabras de la figura 18. Dado que este sistema de recomendación complementa las

métricas generales con datos textuales, las sugerencias que propone son diferentes a las del anterior, pero brindan una idea más precisa de las probabilidades de apoyo a un proyecto según su contenido y la categoría en la que se incluye (ver figura 19).

Esta misma comprobación se puede realizar seleccionando, de manera aleatoria, un proyecto del conjunto de datos de Kickstarter. De igual manera que con el ejemplo de Ideame, el sistema basado en métricas devuelve un número de recomendaciones cuya probabilidad se sustenta en la correlación de los indicadores numéricos resaltados con una coloración rojiza (ver figura 20). Es interesante notar que para el sistema basado en contenido es indiferente si los proyectos sugeridos han sido o no exitosos (ver figura 21), lo que implica que en este caso es la similitud de los contenidos textuales (ver figura 22) la que define la probabilidad de que una campaña sea o no apoyada. Cabe recalcar que ambos sistemas requieren de diversos ajustes y mejoras, pero se espera que su desarrollo ulterior permita a los creadores y creadoras obtener información relevante para el éxito de sus proyectos.

Figura 18. Nube de palabras de los proyectos sugeridos (conjunto de datos Ideame)

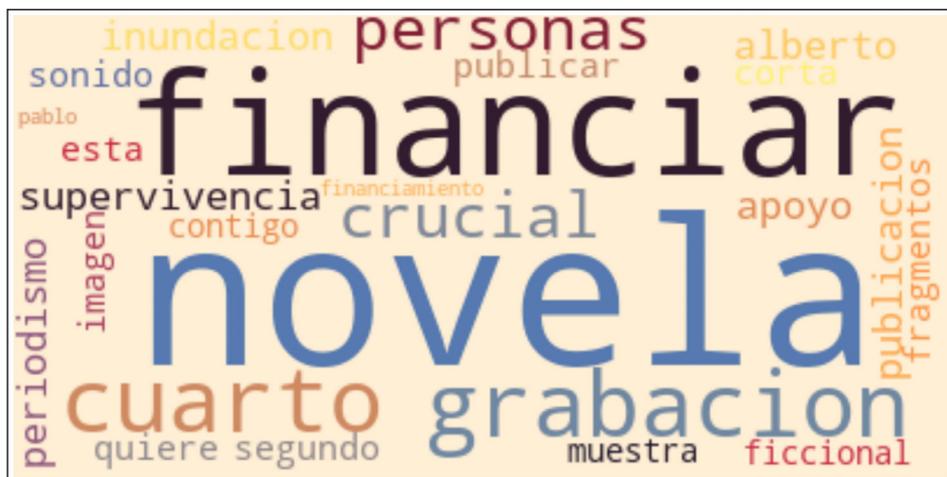


Figura 19. Resultados de recomendación basada en contenidos (conjunto de datos Ideame)

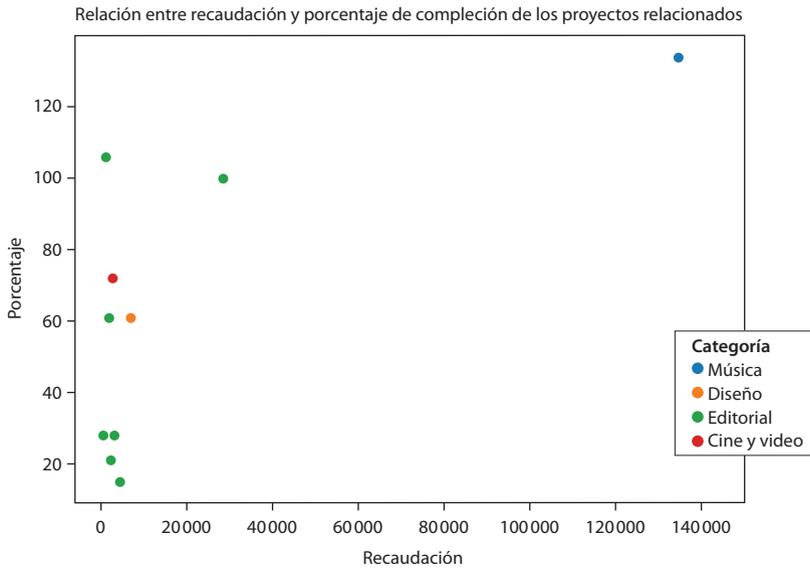


Figura 20. Resultados de recomendación basada en métricas (conjunto de datos Kickstarter)

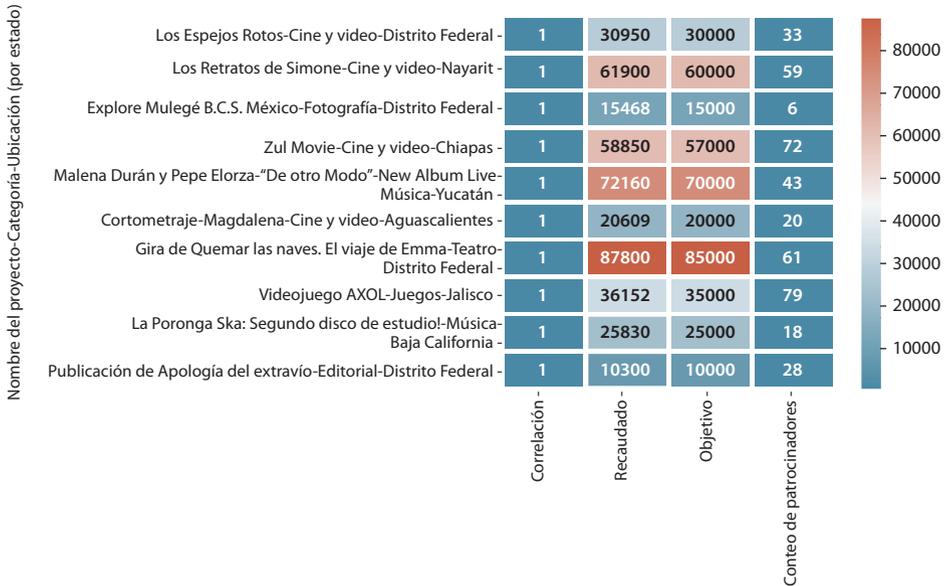


Figura 21. Resultados de recomendación basada en contenidos (conjunto de datos Kickstarter)

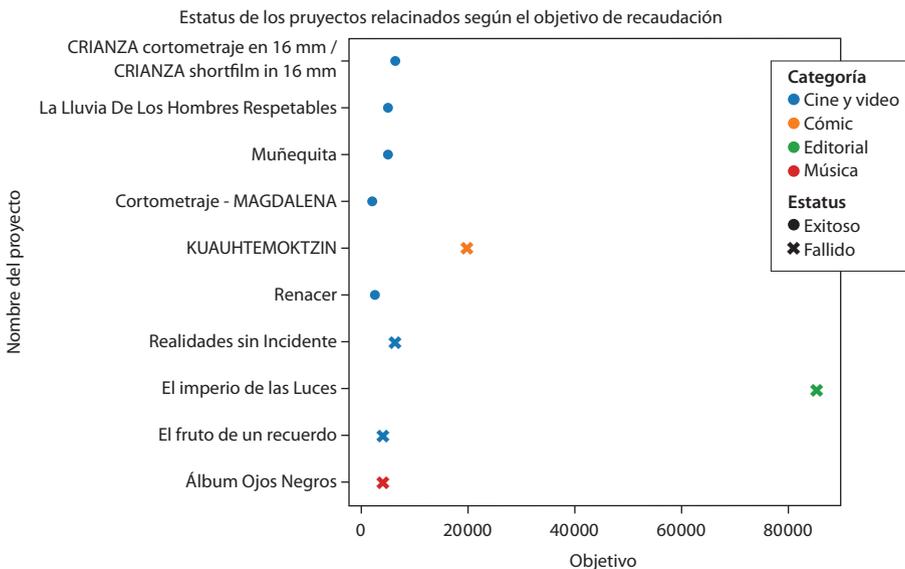
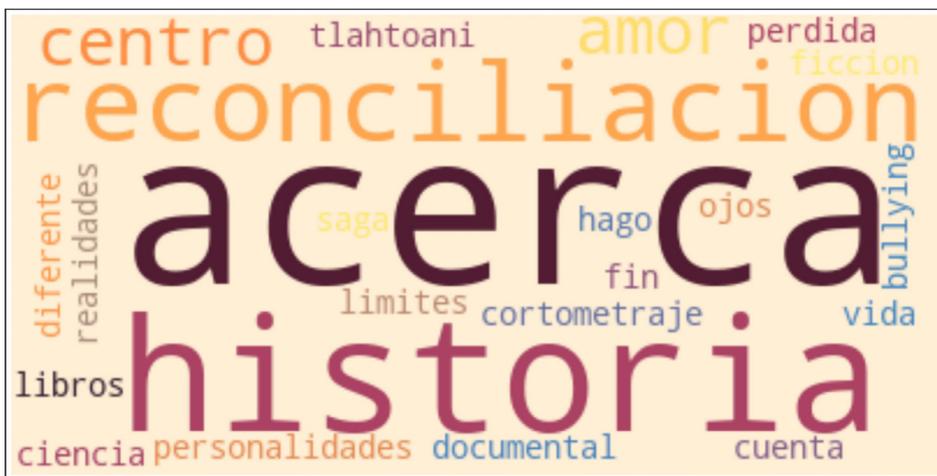


Figura 22. Nube de palabras de los proyectos sugeridos (conjunto de datos Kickstarter)



Aunque aquí concluye este capítulo, el estudio que fue desarrollado en estas páginas puede ser continuado desde diversas ópticas y con diferentes finalidades. En este caso específico, el propósito de esta investigación es el de brindar a creadores y creadoras, gestores y promotores culturales y creativos, información con la que puedan tomar decisiones basadas en datos y algoritmos que conduzcan a una planeación orientada al desarrollo y promoción exitosa de sus proyectos culturales y creativos; no obstante, los sistemas aquí desarrollados pueden implementarse tanto para elaborar un diagnóstico de la producción cultural y creativa en México y en América Latina, lo mismo que para la toma de decisiones estratégicas (Gander, 2017) en torno a la creación y difusión de proyectos o campañas independientes, así como para la captación de colaboradores o patrocinadores en las plataformas digitales de micromecenazgo.



## 2. El microrrelato digital: el caso de Twitter

### 2.1. Microblogging y ficción en red

El progreso de la humanidad, en términos sociohistóricos, pudiera examinarse desde dos ópticas que son, en gran medida, complementarias. Por una parte, el desarrollo, diseminación y preservación de las ideas, valores, obras, tradiciones, costumbres, etc. de un pueblo o comunidad se debe en gran medida a la capacidad, individual y colectiva de establecer vínculos o canales de transmisión y recepción de mensajes; desde las culturas preclásicas hasta la actual sociedad datificada, los seres humanos construimos, perfeccionamos y empleamos diversos canales o medios de comunicación mediante los cuales no sólo compartimos información física o digital, sino que además intercambiamos percepciones, valoraciones, concepciones del mundo y de nuestros semejantes, en suma, nos comunicamos.

Por otra parte, la historia cultural de las sociedades humanas, esto es, la evolución de todos los aspectos simbólicos que las caracterizan es la historia del surgimiento, transformación o desaparición de las diferentes prácticas mediante las cuales se concretan o materializan la imaginación, la creatividad y el saber humano; en términos sencillos, la historia cultural se articula en torno de los medios y modos de producción simbólica de un pueblo o sociedad particular. Desde esta perspectiva, una misiva, una nota periodística, un correo electrónico, un tweet, un cuadro pictórico, la ejecución de una pieza musical o teatral, un baile, un poema son a la vez canales o medios comunicativos (McLuhan, 2007) en tanto difunden y conservan

mensajes de índole diferente —apelativa, referencial, estética, etc.— lo mismo que prácticas de producción simbólica.

De acuerdo con lo anterior, no sería del todo erróneo afirmar que existe una correlación fundamental entre los medios comunicativos y las formas de producción simbólica, de manera parecida a como se sostuvo en el capítulo anterior el vínculo sustancial entre las prácticas culturales y las tecnologías informáticas actuales (Goriunova, 2012). Así pues, se pueden trazar diversas interacciones entre los medios de producción material de significado, o medios de comunicación como tales (Briggs y Burke, 2005), y las prácticas culturales en las que se agrupan las artes y otras expresiones creativas: por ejemplo, la imprenta prepara el terreno no solamente para el surgimiento ulterior de la industria editorial (McLuhan, 2007; Zallo, 1992), sino también para la conformación del campo intelectual específico de la literatura, operante con una autonomía relativa según sus propias reglas y jerarquías (Bourdieu, 2002), lo mismo que los sistemas de transporte industriales propiciaron la exportación y venta de obras de la plástica, convirtiéndose en uno de los factores que favorecieron la configuración del mercado artístico actual (Moulin, 2012).

De manera semejante, hoy en día el internet, que podría considerarse como la red de redes y, por ende, como la tecnología comunicacional de mayor alcance e impacto (Mattelart, 2002), posibilita no sólo el intercambio de datos e información textual, visual, sonora, audiovisual —en suma, de naturaleza multimedia (Creeber, 2009; Jenkins, 2008)—, sino que además propicia la formulación de nuevas prácticas creativas, tales como el hipertexto, o hipermedia narrativo (Landow, 2008), las prácticas artísticas intermediales (Elleström, 2010), las narrativas transmedia (Scolari, 2013) y otras formas expresivas autóctonas de los entornos digitales, basadas fuertemente también en la cultura de la participación, la inteligencia colectiva y el *fandom*, o cultura del aficionado (Jenkins, 2006), como lo son las *wikias*, el *fanart*, la *fanfiction*, entre otras más.

A propósito de todo esto, es pertinente recordar que uno de los rasgos distintivos de la sociedad de la datificación es la de la generación e intercambio constantes de conjuntos de datos masivos (Ontiveros Baeza y López Sabater, 2018); según la terminología de la ciencia de datos, lo mismo que de la analítica cultural (Manovich, 2020), la mayor parte de la información

existente en internet se presenta de forma no estructurada, esto es, sin una estructura u organización intrínseca o *a priori* (Nelli, 2018); una vez que son procesados y ordenados en un formato tabular, los datos —ahora denominados como estructurados— pueden ser examinados mediante herramientas de minería de datos y extracción de conocimiento y empleados ya sea en el ámbito empresarial o en el académico (García Herrero *et al.*, 2018). Con esta breve digresión lo que se desea aclarar es que textos, imágenes, música, videos y otros datos culturales digitales se presentan usualmente sin una organización o estructura previa que facilite su estudio.

Dado que el objeto de estudio de este capítulo en particular es el microrrelato digital en la plataforma de microblogging de Twitter, conviene explicar, a continuación, los conceptos de texto y de microrrelato. En lo que respecta al primer concepto, el vocablo de texto remite a tres acepciones diferentes que varían según su nivel de universalidad o generalidad, esto es, de la extensión de objetos que comprende. Desde una perspectiva semiótica, un texto es un artefacto cuya significación radica en un conjunto de códigos compartidos entre sujetos (Szurmuk y Mckee Irwin, 2009); así pues, un texto puede ser un mensaje verbal, una obra narrativa escrita, una pieza escultórica, una película o un video tutorial, en tanto son objetos que pueden ser comprendidos mediante normas, convenciones y pautas —tácitas o explícitas— referentes a la comunicación verbal, la escrita, la visual y la audiovisual, respectivamente.

Una segunda acepción del término reuniría bajo la etiqueta de texto a todas las formas de documentos conformados por caracteres pertenecientes a un sistema de escritura (Altamirano y Sarlo, 1990); según esta óptica, un texto puede ser igualmente un escrito histórico, un testimonio narrado, una crónica o un conjunto de anales, lo mismo que otras formas de escritura, como la biografía, la hagiografía, la narrativa literaria y el periodismo. En la cultura digital abundan los datos textuales, por ejemplo, un tweet en Twitter, un estado o publicación en Facebook, un marcador textual en Reddit, un relato en Tumblr, un mensaje en WhatsApp o en Telegram son formas de escritura cuyo sentido es asequible en tanto los usuarios posean lo mismo el conocimiento de las convenciones del medio como las capacidades para expresarse empleando los códigos propios de la plataforma, como sería el caso del uso de *emojis*, memes y *stickers*, en los sistemas de

mensajería instantánea, como WhatsApp, o de los hilos y el retweet, en Twitter.

La tercera acepción de texto, que corresponde a la tradición de los estudios literarios y filológicos (Szurmuk y Mckee Irwin, 2009), es la que se refiere a una clase muy especial de constructos discursivos, que son los que incluyen a los relatos de ficción como una forma especializada de escritura. Es importante mencionar que el concepto de ficción es objeto de numerosas teorizaciones, siendo la filosofía y la teoría literaria las disciplinas destacadas en la reflexión del tema. El estudio de la ficción, cabe agregar, posibilita una comprensión más profunda de las dinámicas con las que percibimos y representamos —o construimos— nuestro entorno tanto en términos conceptuales como imaginativos, sensoriales y emocionales (Martínez Bonati, 2001); en otros términos, la ficción es un lenguaje que se refiere a la realidad redescribiéndola; más adelante se retomará este tema, aunque de una manera muy general, para precisar las características de los tweets ficcionales.

Conforme a lo anterior, con la noción de texto que se emplea en este estudio se intenta compaginar las acepciones previamente comentadas, de manera tal que por texto se puede entender tanto la materialidad misma por la cual se expresa un mensaje (un libro, un panfleto, una pantalla), así como todo constructo discursivo, esto es, todo acto comunicativo expresado mediante caracteres cuya comprensión depende, primero, de que el lector cuente con el conocimiento formal de un sistema signico en particular (como lo es el alfabeto o abecedario de un idioma) y, segundo, de que participe de las convenciones genéricas y del contexto de enunciación de un texto (Cros, 1986), es decir, de que posea las competencias para distinguir una nota periodística, una frase irónica o un relato ficcional.

Con respecto al microrrelato, es éste una forma textual que surgiría y se consolidaría durante las primeras vanguardias como una apuesta por una poética de la brevedad y una estética minimalista (Lagmanovich, 2006); no obstante, algunas influencias de este género podrían rastrearse en la literatura aforística, como en los *Pensamientos*, de Blaise Pascal, o en los aforismos de Georg Christoph Lichtenberg, que tendían a expresar filosofemas a través de parábolas breves; de igual manera, las ficciones breves de Franz Kafka, de Macedonio Fernández, de Jorge Luis Borges, de Juan José Arreola, de Luisa Valenzuela (por mencionar algunos autores), los poemas narra-

tivizados de Charles Baudelaire, lo mismo que los caligramas de Guillaume Apollinaire o el género poético del haikú pueden considerarse sin duda como antecedentes de la microficción. En Hispanoamérica el microrrelato ha tenido un singular atractivo, posiblemente por las posibilidades de exploración y autorreflexión que esta forma literaria ofrece a escritoras y escritores (Siles, 2007).

Pese al entusiasmo con el que es acogida en la literatura contemporánea, la microficción incita a reflexionar en torno al carácter mismo del ejercicio de escritura ficcional, lo mismo que sobre los contactos e hibridaciones que existen entre las diversas prácticas artísticas, así como sobre su clasificación como género y su autonomía como expresión literaria. Sobre esto último, Lagmanovich (2006) sostiene que el microrrelato no es resultado de las intersecciones con otros géneros literarios o formas artísticas, argumento que mantiene con miras a sustentar desde su perspectiva teórica la independencia de la microficción; no obstante, varios estudiosos del fenómeno artístico y literario afirman que no existen formas puras y que, al contrario, los géneros y las prácticas creativas en general son producto de múltiples relaciones intertextuales e intermediales (Arvidson, Askander, Bruhn y Fuhrer, 2007; Elleström, 2010).

Por otra parte, el microrrelato plantea una serie de desafíos teóricos, de entre los cuales se destacan el de la delimitación del género, las formas y sub-formas que engloba el vocablo y su condición o estatus estético como artefacto literario (Siles, 2007) o, en otras palabras, qué características posee un texto para ser considerado como microficción, qué tipos de textos pueden incluirse en este género y qué aspectos determinan su valor literario y su potencial estético (Barthes, Lefebvre y Goldmann, 1969). Respecto a la cuestión del género, los atributos del microrrelato serían la brevedad, la narratividad y la simplicidad sintáctica (Lagmanovich, 2006); en lo que concierne a la brevedad, ésta no sería una característica cuantificable o medible, es decir, un microrrelato no se constriñe a una cantidad de palabras—o, en el caso de las microficciones de Twitter, a un número de caracteres—, antes bien, la brevedad alude a la resolución concisa del relato (Álamo Felices, 2010).

La narratividad en el microrrelato, por otra parte, se refiere a la existencia de un núcleo argumental básico, una anécdota, suceso o aconteci-

miento inicial, y a un incidente que, de manera súbita, introduce una transformación y que conduce a un desenlace, ya sea puntual o abierto (Lagmanovich, 2006); con respecto a la simplicidad sintáctica, ésta consiste en evitar las construcciones gramaticales complejas —como lo sería la perífrasis verbal recargada o la concatenación excesiva de oraciones subordinadas— en pos de una estructura sencilla, precisa, en la que el argumento se exponga sin mayor dilación, sin adornos ni barroquismos innecesarios (Álamo Felices, 2010). El género del microrrelato incluiría, pues, a todos aquellos textos de ficción que se caracterizan por tener una anécdota mínima, expresada de manera concisa y con una estructura lo más simple posible.

En el caso específico de este estudio, se parte del supuesto de que toda forma textual que cumpla con las condiciones y características previamente descritas de la brevedad, la narratividad y la simplicidad sintáctica, forma parte del corpus de análisis de este estudio, independientemente del medio o soporte a través del cual se difunde o se distribuye. En este sentido, algunos de los textos que producen los usuarios de sitios de microblogging como Twitter, Reddit o Tumblr pueden considerarse como microrrelatos (Bocanegra Barbecho y García López, 2017) si reúnen las propiedades atribuibles al género. Cabe señalar que el corpus aquí examinado se conforma solamente por textos recabados de Twitter, dado que esta plataforma cuenta con una comunidad hispanohablante más activa.

Sobre la cuestión relativa al valor literario y estético del microrrelato, concretamente de las microficciones en redes sociales de microblogging, como es el caso de este estudio, el que un texto se considere como literario depende, en gran medida, de una serie de regulaciones y validaciones institucionales que no siempre se vinculan con las capacidades creativas o deseos del productor ni con las propiedades específicas o diferenciadoras del producto textual (Escarpit, 1980), de igual manera que una práctica artística se valida por las diversas interacciones sociales que operan entre creadores, intermediarios y audiencias (García Canclini, 2014; Moulin, 2012). En lo que respecta a los microrrelatos que conforman el corpus de este estudio, el problema de su carácter literario o de su valor estético radica en el hecho de que son formas de producción textual que no están sujetas a un régimen literario canónico: no son microficciones producidas por escritoras o escritores consagrados ni representan a una casa editorial de renombre.

La interrogante sobre los criterios por medio de los cuales se pudiera determinar el valor estético o la condición literaria de un tweet a través del cual se relata una historia mínima, perfila la cuestión central de este estudio, que es la de examinar las características diferenciadoras de microtextos de ficción, o microficciones como tales, empleando herramientas y técnicas de procesamiento de lenguaje natural (NLP), con las cuales se identifiquen patrones sintácticos, estilísticos, semánticos y temáticos que conduzcan al desarrollo e implementación de un sistema de clasificación automatizado de textos, basado en aprendizaje máquina supervisado (*Supervised Machine Learning*) (Albon, 2018), que posibilite la catalogación de miles de textos —independiente de sus respectivas extensiones— en ficcionales, noticiosos, de hechos o de opinión, etcétera.

Una de las dificultades teórico-metodológicas que surgen de este estudio es la de si existen características o propiedades que permitan distinguir entre textos —o microtextos, como en este caso— de ficción, de opinión, académicos, etc., observables a través del lenguaje escrito o, en otros términos, ¿es posible identificar los patrones léxicos, sintácticos y semánticos de un relato o microrrelato de ficción? Como se advirtió anteriormente, se retoma aquí el tema de la ficción, no obstante, intentar elaborar un concepto estable de ésta sería una tarea que, al involucrar diferentes abordajes disciplinares, sobrepasaría el propósito de este capítulo, que es el de emplear técnicas de procesamiento de lenguaje natural para identificar las características y los patrones de la escritura de microficciones en Twitter a través de un corpus extenso de tweets. En lugar de ello, se ofrece, a continuación, un compendio panorámico de los enfoques teóricos con los que se ha explorado la ficcionalidad (Garrido Domínguez, 1997), mismo que servirá para delimitar y especificar las características del microrrelato ficcional digital:

(a) *Enfoque ontológico*: la ficción literaria no solamente crea objetos, sino conjuntos de objetos, es decir, los entes ficcionales no existen de manera aislada, sino que se encuentran sujetos a un *aquí* y a un *ahora*, habitan y cohabitan en *mundos posibles*. La noción remite, obviamente, a Leibniz, pero su uso actual dista mucho del de las especulaciones teológico-metafísicas del filósofo alemán. El más consistente desarrollo de este concepto en la teoría literaria se debe a Lubomir Doležel, para quien estos universos, pro-

ducidos por la actividad imaginario-poética, coexisten de diferentes maneras con el mundo factual o real.

(b) *Enfoque pragmático*: la ficción literaria es esencialmente un acto comunicativo, aunque con marcadas peculiaridades. Si bien la base de este enfoque es la de la teoría de los actos de habla, diversos autores discrepan de las conclusiones expuestas por Searle, que niega a los actos de habla ficcionales el carácter de autenticidad y validez. Antes bien, los actos de habla ficcionales no aspiran a la autenticación referencial de la comunicación plenamente constituida en cuanto práctica cotidiana, sino a la validación según determinadas convenciones culturales establecidas por una institución o un tipo de conducta a la que Schmidt (1997) denomina LITERATURA.

(c) *Enfoque antropológico-imaginario*: la ficción literaria tiene su fundamento en el impulso humano de traducir los fenómenos que acaecen en el mundo empírico a símbolos que se ajusten a la experiencia mediadora con la realidad externa. El texto ficcional mantiene una conexión con el mundo real en cuanto devela las estructuras simbólicas por las que la psique construye las vivencias. Entre los autores representativos de esta perspectiva se pueden hallar a Gilbert Durand, quien propone una tipología basada en la influencia del mito, y a Wolfgang Iser, quien postula la estructura del doble significado, de ocultación y revelación, de la ficcionalidad.

El enfoque por el que se opta en este capítulo es el pragmático, lo que implica que la condición de “fictivo” o de ficcionalidad de un texto radica en lo que Schmidt denomina como *convención estética*, que consiste en la expectativa de que los textos ficcionales cumplan con las siguientes funciones: (a) “restar énfasis a la convención factual” (Schmidt, 1997, p. 223), es decir, los textos de ficción se sustraen a la obligación de presentar tal o cual afirmación respecto a un orto-modelo de mundo como verdadera o real; (b) ofrecer señales claras y apropiadas para la recepción de un texto como específicamente literario, y (c) “seleccionar como marco de referencia para las expresiones referenciales que se encuentran en el texto no sólo el modelo de realidad socialmente establecido, sino otros marcos de referencia” (Schmidt, 1997, p. 224).

Los tweets ficcionales con los que se conformó el corpus de este estudio cubren las características anteriormente comentadas de la brevedad, la narratividad y la simplicidad sintáctica, por lo que se asume que pertenecen

al género denominado como microrrelato (Lagmanovich, 2006), además de cumplir con las condiciones de la convención estética postulada por Schmidt, todo lo cual se podrá corroborar más adelante, en el abordaje analítico del corpus. Cabe agregar que, antes de proceder al análisis del corpus, así como al desarrollo del sistema de clasificación automatizada de textos, es pertinente primero explicitar los criterios y estrategias mediante los cuales se elaboró el corpus de trabajo, además de describir en qué consiste el NLP y qué técnicas y métodos lo constituyen.

## 2.2. Técnicas y aplicaciones del procesamiento de lenguaje natural

Si bien el concepto de NLP no es un tópico habitual en el día a día, sus técnicas, así como los productos y aplicaciones que se derivan de éste, tienen un uso cotidiano frecuente al grado de que se han convertido en parte de la rutina comunicativa humana. Cuando empleamos Google Traductor para conocer el significado de alguna frase en algún idioma que desconocemos o bien la compleción automática de palabras en el caso de la mensajería instantánea, lo mismo que las sugerencias de búsqueda que aparecen en nuestros navegadores y las recomendaciones de eventos o productos en redes sociales, estamos ante algunos de los usos y aplicaciones resultantes del tratamiento computarizado del lenguaje humano.

Algunas prácticas más complejas del NLP involucran herramientas para elaborar resúmenes o sintetizar grandes volúmenes de texto o para extraer los tópicos y las palabras clave de un texto, lo mismo que para clasificar documentos según categorías, que es la aplicación resultante de este estudio. Además, las plataformas de detección de plagio, como Plag Scan o Antiplagiarism operan mediante técnicas y algoritmos propios del NLP, como sería la comparación de similitudes entre textos y la detección de frecuencia de palabras, entre otras más. Al tratarse de un subcampo derivado de la Inteligencia Artificial, así como del aprendizaje máquina, también se han hecho adelantos significativos en el desarrollo de agentes o asistentes conversacionales, como Siri, Cortana y Google Assistant para sistemas operativos OS, Windows y Android, respectivamente.

De acuerdo con varios especialistas (Beysolow II, 2018; Dwivedi, 2020; Kulkarni y Shivananda, 2019), en el subcampo del NLP convergen disciplinas tales como la lingüística, las ciencias computacionales y la inteligencia artificial con el propósito común de analizar los llamados “lenguajes naturales”, que son sistemas sígnicos desarrollados mediante las interacciones entre hablantes para expresarse y comunicarse, y que se diferencian de los lenguajes formales, que son los sistemas artificiales empleados, por ejemplo, en la lógica y las matemáticas, para representar un aspecto específico de un fenómeno de manera abstracta (Lyons, 1984). El desarrollo del NLP se ha intensificado de manera exponencial gracias a las investigaciones de Alan Turing, que derivarían posteriormente en el aprendizaje máquina (Lane, Cole y Max Hapke, 2019).

A partir de las décadas de 1950 y 1960, el ámbito del NLP se divide en dos enfoques, uno liderado por lingüistas y el otro, por estadistas y matemáticos; los primeros abogaban por una aproximación simbólica al estudio del lenguaje humano, y sus métodos de análisis se enfocaban en el reconocimiento de patrones, mientras que los segundos proponían un abordaje estocástico, centrando sus estudios en la aplicación de algoritmos de agrupamiento (*clustering*) (Beysolow II, 2018); en la actualidad, pareciera que ambos enfoques operaran de manera conjunta, pues el subcampo del NLP ha ampliado sus objetivos más allá del análisis lingüístico computacional y del tratamiento estadístico del lenguaje para llegar a una comprensión contextual de las frases y los discursos, en lo que se denomina entendimiento del lenguaje natural (NLU, *Natural Language Understanding*) (Beysolow II, 2018; Vasiliev, 2020).

Para brindar una visión más detallada del NLP, así como de su empleo en el estudio de las artes y la cultura en general (Burdick, 2012; Manovich, 2020), es pertinente distinguir entre las tareas o técnicas usuales relacionadas con este subcampo del aprendizaje máquina y las aplicaciones que se derivan de éste (Kulkarni y Shivananda, 2019); en lo que respecta al primer rubro, se entiende por tarea o técnica de NLP todos aquellos pasos o actividades para el preprocesado y reducción de la dimensionalidad de uno o varios corpus lingüísticos (Dwivedi, 2020); por otra parte, las aplicaciones del NLP se refieren al empleo de un conocimiento derivado de este ámbito para solucionar necesidades o resolver problemas relativos a los usos y/o comprensión del lenguaje humano.

Una de las tareas o técnicas fundamentales del procesamiento de lenguaje natural es la recolección automatizada de texto (Kulkarni y Shivananda, 2019), que consiste en, primero, identificar una fuente de información útil para los propósitos que se plantean para el proyecto en curso, para, segundo, extraer el contenido textual implementando un flujo de trabajo específico; una vez extraída tal información se procede, por último, a organizarla según variables, metadatos u otros identificadores (Science, 2020). Las fuentes de información a las que se puede recurrir en un proyecto de NLP son variadas, pero se pueden clasificar en (a) las que son provistas por los clientes o interesados (bases de datos, almacenamiento en la nube u otros archivos brindados por un tercero), (b) las de acceso gratuito —limitado o no—, tales como repositorios de datos abiertos gubernamentales o de la iniciativa privada, API, tema que se abordó en el capítulo anterior, o sitios informativos construidos por comunidades, como Wikipedia; (c) mediante el web scraping, o raspado de red, se puede trocar uno o varios sitios web en fuentes de información (Kulkarni y Shivananda, 2019), como se ejemplificó en el capítulo anterior.

Independientemente de la clase de fuentes de información descritas anteriormente, la recolección automatizada de texto se lleva a cabo para extraer datos textuales de diferentes tipos de objetos o documentos, como son archivos de editores de texto (Microsoft Word, o Writer, de Libre Office), archivos PDF, ficheros .csv, objetos JSON para el intercambio de datos (JavaScript Object Notation), páginas basadas en lenguaje de marcado HTML o sitios web con mayor nivel de interactividad (basadas en JavaScript). Conforme a las destrezas y habilidades del investigador, así como del entorno de trabajo con el que se realiza el proyecto, la recopilación de texto se puede personalizar a tal grado que incluso se pueden extraer párrafos, páginas, frases, palabras o partículas gramaticales específicas (Sarkar, 2019).

Una vez que se han recabado los datos textuales, la siguiente tarea es el preprocesamiento, que comprende actividades de limpieza, estandarización y transformación del texto para garantizar su legibilidad y facilitar el desarrollo de la aplicación (Verma, 2020). Durante la fase de preprocesamiento, el texto se convierte a minúsculas, se remueven los signos de puntuación y las palabras vacías —o *stopwords*—, se eliminan símbolos o cadenas de caracteres no esenciales, como *emojis*, emoticonos o enlaces, se estandariza el

texto, esto es, se sustituyen palabras mal pronunciadas o incompletas o se traducen secuencias en otros idiomas diferentes al de interés para el proyecto, se *tokeniza* el texto, es decir, se divide en unidades semánticas mínimas, se extraen las raíces morfológicas de los vocablos relevantes —en inglés, el proceso se denomina como *stemming*—, y se ubican los “lemas” o palabras que, por su significado, abarcan a otras —*lemmatization*, en inglés— (Dwivedi, 2020; Kulkarni y Shivananda, 2019).

Los datos textuales no poseen una estructura tabular inherente, es decir, no se presentan organizados según un orden o clasificación inteligible; antes bien, uno de los objetivos de todo proyecto de NLP es ordenar uno o más corpus de textos de manera que sean comprensibles para el usuario humano, a la vez que puedan ser codificados y decodificados por la máquina (Sarkar, 2019); para ello, se requiere transformar los datos textuales en “atributos”, esto es, en un tipo de formato numérico que facilita su exploración y análisis (Kulkarni y Shivananda, 2019). Existen diferentes técnicas de transformación textual, tales como la *One Hot Encoding*, la cual consiste en convertir variables categóricas ordinales en números enteros que van del cero hasta cubrir todos los valores (Nelli, 2018). Es importante señalar que esta técnica es más eficaz para valores textuales singulares, como palabras específicas.

Otra técnica de transformación de datos textuales a atributos es la de la vectorización por conteo, que implica contabilizar la frecuencia con la que una palabra en particular aparece en el texto (Beysolow II, 2018); una técnica más general que las previamente comentadas es la de los n-gramas, que consiste en dividir las frases u oraciones en segmentos de dos (bigramas), tres (trigramas) o más palabras recurrentes en el corpus que faciliten el modelado de los textos. La matriz de co-ocurrencia, a diferencia de la vectorización por conteo y los n-gramas, permite agrupar oraciones completas según su frecuencia en el texto procesado (Kulkarni y Shivananda, 2019). Una de las técnicas más avanzadas en la actualidad es la de la vectorización por segmentos, que convierte sobre la marcha las cadenas de caracteres en índices numéricos (Sarkar, 2019).

La frecuencia de término-frecuencia inversa de documento (*TF-IDF*, *Term Frequency-Inverse Document Frequency*) calcula la relevancia de una palabra en una frase en función de la longitud del texto o documento en el que aparece (Vasiliev, 2020). Esta técnica es básica para el NLU en tanto

posibilita una comprensión contextual del texto examinado. Por último, las incrustaciones de palabras (o *Word Embeddings*) van un paso adelante de la técnica del TF-IDF, en tanto consiste en un mapeo general de todas las palabras que aparecen en cada uno de los documentos que conforman el corpus, para representarlas numéricamente en función de las relaciones que tienen unas con otras, de manera individual, lo mismo que como oraciones, según su contexto en la totalidad del corpus (Vasiliev, 2020; Verma, 2020).

Las aplicaciones del procesamiento de lenguaje natural varían según su complejidad, esto es, conforme a las diferentes técnicas o tareas que incorporen, así como al volumen de datos con el que se construye el modelo y la exactitud o refinamiento del producto o servicio final esperado (Beysolow II, 2018). A continuación se enlistan algunas de las aplicaciones elementales del NLP: (a) el etiquetado de partes del discurso (*POS tagging: Part of Speech*), que consiste en identificar la función gramatical de una palabra, según su posición en la frase y su contexto; (b) la detección de similitudes entre textos, mediante la cual se determina la proximidad léxica y semántica entre dos textos (Sieg, 2019); (c) el reconocimiento de entidades nombradas (NER, *Named Entity Recognition*), que son las palabras o frases con un referente “real”; (d) la extracción de información, que consiste en la selección y filtrado de partes de un texto, acorde a parámetros determinados; y (e) la síntesis o resumen de textos, que es la reducción de un texto extenso, mientras se conserva su información esencial (Gonçalves, 2020; Kulkarni y Shivnanda, 2019).

De entre las aplicaciones que pudieran considerarse como intermedias, por su grado de complejidad, está el modelado de tópicos, que se basa en técnicas heurísticas, tales como la TF-IDF o las incrustaciones de palabras que buscan descubrir temas recurrentes en un corpus de textos (Lane *et al.*, 2019); a este tipo de aplicaciones se suma la clasificación de textos, que consiste en organizar un conjunto extenso de documentos textuales en categorías específicas según sus similitudes, atributos o tópicos recurrentes (Vasiliev, 2020). Precisamente, la aplicación resultante de este estudio es un sistema de clasificación textual multiclase, el cual, a diferencia de las clasificaciones binarias —que son las más comunes—, permite catalogar los textos en clases más específicas (Li, 2020). El análisis de sentimientos es otra aplicación de media a alta complejidad del NLP, que parte de la clasificación

de textos, pero añade a ésta una ponderación de la polaridad de ciertas palabras —si remiten a emociones o sentimientos positivos, negativos o neutros— (Mathur, 2019).

Finalmente, las aplicaciones de mayor complejidad incluyen la detección y traducción de lenguajes que, como su nombre lo indica, consiste en identificar el idioma de un texto para después traducirlo a otro idioma según las necesidades del usuario (Kumar, 2021). El reconocimiento de voz y su conversión a texto, así como la transformación de texto a voz, son aplicaciones que permiten al usuario interactuar de una manera más directa con la máquina, y son la base sobre la cual se desarrollan actualmente los llamados —chatbots—, o asistentes conversacionales, que son programas informáticos que desempeñan una serie de actividades que van desde responder preguntas frecuentes de usuarios, ofrecer atención a clientes en cualquier momento, realizar encuestas e incluso brindar apoyo psicológico básico,<sup>9</sup> seguimiento de información financiera y algunas tareas educativas elementales —como la toma de asistencia— (Raj, 2019).

Tras este sucinto repaso de las técnicas y aplicaciones del NLP, es pertinente esclarecer los criterios y técnicas mediante las cuales se conformó el corpus para este estudio. El corpus, como se ha indicado previamente, se conformó a partir de la recolección de una gran cantidad de tweets, a los cuales se tuvo acceso mediante la API, que proporciona Twitter (Kulkarni y Shivananda, 2019). En un principio se extrajeron solamente los textos que coincidieron con la búsqueda del término de “microficción” (ver tabla 7); no obstante, para desarrollar un sistema de clasificación verdaderamente eficiente se requería diferenciar los tweets ficticiales de otros tipos de documentos textuales, por lo que se extrajeron también tweets de noticieros (ver tabla 8) y de frases y reflexiones (ver tabla 9) mediante la función de búsqueda integrada a la misma API de Twitter. Con la aplicación de determinadas técnicas de NLP, en el siguiente apartado se analiza si los tweets ficticiales cumplen con las condiciones de la brevedad, la narratividad, la simplicidad sintáctica y la convención estética que se comentaron en el primer apartado de este capítulo.

---

<sup>9</sup> Un ejemplo clásico es Eliza, desarrollada entre 1964 y 1966 por Joseph Weizenbaum, un profesor reputado del Massachusetts Institute of Technology (MIT).

Tabla 7. Conjunto de tweets de relatos breves ficcionales

	<i>Texto</i>	<i>Usuario</i>	<i>Fuente</i>	<i>Conteo de retweets</i>	<i>Conteo de favoritos</i>
0	Sonrió a la mujer que se sentó junto a él en el parque. Fue sutil pero ella le recordó que tenía alzheimer y que ya estaban casados.	Narratorio	TweetDeck	120	320
1	Le ataron en medio del campo a unos tablones y le dejaron a su suerte, mecido por el viento. Lo peor es la soledad, pensaba el espantapájaros	Narratorio	TweetDeck	28	70
2	“—La culpa es del corazón que me donaron, era de un asesino. Los policías se miraron sorprendidos. Fue suficiente para que sacara el arma y los liquidara. —Idiotas— mira que creer semejante tontería... #Panicosiniestro”	Narratorio	TweetDeck	13	29
3	“—Dicen que si te tatúas su firma él se queda con tu alma. —Jajaja. Qué cosas se cree la gente, ¿Verdad? —Pues sí. Pero es el tatuaje más popular en el estudio, te lo hago al 50% por ser tú. —Venga, va. Al tatuador le brillaron los ojos de un sospechoso color rojo. #PanicoSiniestro”	Narratorio	TweetDeck	13	27
4	Delicioso, pensó sorprendido mientras masticaba, y se rió al recordar que casi se entrega a la poli porque no sabía qué hacer con el cadáver	Narratorio	TweetDeck	12	23
5	La madre superiora no dudaba en aplicar severos correctivos. Tras años de desapariciones en el convento, las monjas descubrieron que la madre superiora había muerto hacía medio siglo. #panicosiniestro	Narratorio	TweetDeck	12	20

Tabla 8. Conjunto de tweets de noticias

	<i>Texto</i>	<i>Usuario</i>	<i>Fuente</i>	<i>Conteo de retweets</i>	<i>Conteo de favoritos</i>
0	Responde la DEA a Gertz: tenemos 95% de éxito en casos; la FGR, 5% <a href="https://t.co/JLkFI0dLN">https://t.co/JLkFI0dLN</a> <a href="https://t.co/7z2i9Gp1Uy">https://t.co/7z2i9Gp1Uy</a>	AristeguiOnline	Hootsuite Inc.	4484	14389
1	Por un decreto presidencial, Cienfuegos es asesor de la Sedena en actual gobierno <a href="https://t.co/kSNV6ifK6g">https://t.co/kSNV6ifK6g</a> <a href="https://t.co/VLc9MgmBJI">https://t.co/VLc9MgmBJI</a>	AristeguiOnline	Hootsuite Inc.	4426	9122

2	Felipe Calderón y Fidel Herrera ocultaron la violación a Ernestina Ascencio: perito <a href="https://t.co/cptLG3t1o2">https://t.co/cptLG3t1o2</a> <a href="https://t.co/joKe1Mblok">https://t.co/joKe1Mblok</a>	AristeguiOnline	Hootsuite Inc.	1360	2496
3	Confirma ISSSTE robo de vacunas en Morelos e informa de investigación en curso <a href="https://t.co/NroPFwC9T0">https://t.co/NroPFwC9T0</a> <a href="https://t.co/A1pOPfrFpR">https://t.co/A1pOPfrFpR</a>	AristeguiOnline	Hootsuite Inc.	1163	2410
4	#AristeguiEnVivo   En el momento que regresaron a Cienfuegos a México, yo sabía que no lo iban a procesar. Es un caso de impunidad, pues las pruebas en contra del general son contundentes: Mike Vigil, Ex jefe de operaciones internacionales de la DEA <a href="https://t.co/rD0NgpirnJ">https://t.co/rD0NgpirnJ</a> <a href="https://t.co/cROHKQVAjc">https://t.co/cROHKQVAjc</a>	AristeguiOnline	Twitter Web App	805	1924
5	#Alerta   García Harfuch recibía 200 mil dólares mensuales de Guerreros Unidos: testigo <a href="https://t.co/Z7qy8CUTVt">https://t.co/Z7qy8CUTVt</a> <a href="https://t.co/Ve7zjK2wwo">https://t.co/Ve7zjK2wwo</a>	AristeguiOnline	Hootsuite Inc.	770	1758

Tabla 9. Conjunto de tweets de frases, pensamientos y reflexiones

	<i>Texto</i>	<i>Usuario</i>	<i>Fuente</i>	<i>Conteo de retweets</i>	<i>Conteo de favoritos</i>
0	Que ganas de no escuchar la palabra coronavirus nunca mas en la vida.	SoloReflexiones	TweetDeck	7941	16058
1	Por salud mental a veces es mejor no saber nada.	SoloReflexiones	TweetDeck	5826	13656
2	<a href="https://t.co/U0q113Wxj3">https://t.co/U0q113Wxj3</a>	SoloReflexiones	TweetDeck	5410	13571
3	Somos lo que damos. <a href="https://t.co/A9GbjwoGEw">https://t.co/A9GbjwoGEw</a>	SoloReflexiones	TweetDeck	5302	12762
4	Verte triunfar. <a href="https://t.co/68v6KEJDkm">https://t.co/68v6KEJDkm</a>	SoloReflexiones	TweetDeck	4969	13430
5	"Conozco muy bien mi propia historia, por lo tanto, soy el único que puedo juzgarme, criticarme y aplaudirme cuando yo quiera."	SoloReflexiones	TweetDeck	4949	2464

Una vez que se unieron todos los textos en un solo conjunto de datos, el corpus para este estudio tuvo una extensión de 9000 documentos, distribuidos de manera equivalente en tweets de relatos breves ficcionales, de noticias de actualidad y de reflexiones o pensamientos (ver tabla 10). Cabe agregar que el sistema de clasificación multiclase que se desarrolla a partir de este estudio se considera como una implementación de aprendizaje máquina de tipo supervisado, en tanto las observaciones se adaptan y se agrupan según categorías definidas previamente, a diferencia del aprendizaje máquina de tipo no supervisado, en el que el modelo se construye según las características o atributos que logre identificar de las observaciones mismas del fenómeno (Albon, 2018).

Tabla 10. Muestra del corpus

	<i>Texto</i>	<i>Noticia</i>	<i>Frase/Pensamiento</i>	<i>Microficción</i>
721	García Luna quiere saber quiénes son los testigos en su contra <a href="https://t.co/xeugctB8xP">https://t.co/xeugctB8xP</a> <a href="https://t.co/wMXGqHd7eE">https://t.co/wMXGqHd7eE</a>	1	0	0
6957	“- ¿Ya es lunes? No puede ser. Voy tarde. - ¿Y qué importa? - Hoy comienza la invasión. - Bueno, no se van a ir sin ti, eres el general, quédate un ratito más – susurró. Con un whatsapp el espía logró advertir a su país evitando una masacre.”	0	0	1
5635	Me delata la sonrisa cuando escucho su nombre.	0	1	0
539	Casi la mitad de peruanos rechaza vacunarse contra la Covid-19 <a href="https://t.co/F3dYJbOATS">https://t.co/F3dYJbOATS</a> <a href="https://t.co/dWjpnDjYFN">https://t.co/dWjpnDjYFN</a>	1	0	0
3029	Recuerden lavarse las manos y no meterse en lo que no les importa.	0	1	0
3901	El corazón también se cansa de esperar, de callar y de intentar.	0	1	0

Es importante indicar que, como se observa en la tabla 10, las categorías o columnas en que se organiza el corpus comprenden el texto, aún no preprocesado, de los tweets recabados, así como las clases en las que se agrupan los diferentes documentos, las cuales contienen los valores 0 y 1, dependiendo de si el texto pertenece o no a la categoría. La razón de que los documentos se identifiquen mediante este rango es que el sistema de clasificación desarrollado opera con base en datos lógicos o booleanos, lo que en este caso significa que si al texto se le asigna un valor de 0, equivale a que se le declare como “falso” —el texto *no es* una microficción, por ejemplo—, pero si se le otorga el valor de 1, entonces es “verdadero” —el texto *es* una microficción, continuando con el ejemplo previo—. Si bien la categorización con la que se construyó el corpus puede parecer arbitraria, el análisis que antecede al modelo de clasificación, y que a continuación se presenta, permite observar diferencias particulares entre las tres clases de tweets.

### 2.3. PLN/NLP para el análisis y clasificación de microtextos digitales

El primer paso para el desarrollo del modelo de clasificación textual es el preprocesamiento de los documentos, proceso que, como se indicó ante-

riormente, comprende la limpieza, estandarización y transformación de los textos. En el caso del corpus de este estudio, se llevaron a cabo las actividades de eliminación de signos de puntuación y de palabras vacías (*stopwords*), así como de *emojis*, emoticonos y enlaces a sitios web; enseguida todos los textos fueron cambiados a minúsculas, se tokenizaron —se dividió cada texto en unidades semánticas mínimas— y se realizaron dos comprobaciones para localizar valores duplicados y valores nulos o cadenas de caracteres vacías, tras lo cual se removieron 1203 observaciones de tweets repetidos o cuyo contenido consistía solamente en emoticonos o vínculos a páginas electrónicas. El corpus preprocesado incluye una nueva columna titulada “Texto limpio” (ver tabla 11) y un total de 7797 documentos.

Tabla 11. Muestra del corpus tras preprocesamiento

<i>Texto</i>	<i>Texto limpio</i>	<i>Noticia</i>	<i>Frase/ Pensamiento Microficción</i>		
6323	Para los viajeros del Cuatro Vientos su hazaña no era suficientemente grande, necesitaba un halo de misterio. Acordaron aterrizar en México, vender el avión por piezas y vivir en el anonimato hasta su muerte para asegurarse un buen sitio en la historia. #asilooscuro @AsiloOscuro	viajeros cuatro vientos hazaña suficientemente grande necesitaba halo misterio acordaron aterrizar méxico vender avión piezas vivir anonimato muerte asegurarse buen sitio historia asilooscuro	0	0	1
2473	#AristeguiEnVivo   Hablamos en directo con @Netzai_Sandoval, Titular del Instituto Federal de Defensoría Pública, sobre la reforma judicial que modifica es estatus del instituto <a href="https://t.co/Qma53Hp4rb">https://t.co/Qma53Hp4rb</a> <a href="https://t.co/Kh1wr1LyUg">https://t.co/Kh1wr1LyUg</a>	aristeguienvivo hablamos directo netzai_sandoval titular instituto federal defensoria pública reforma judicial modifica estatus instituto	1	0	0
1885	Debaten @DeniseDresserG y @sabinaberman: ¿deben continuar las mañaneras durante la campaña electoral? <a href="https://t.co/FPn86tQFJP">https://t.co/FPn86tQFJP</a> <a href="https://t.co/h5DV3Ub6jt">https://t.co/h5DV3Ub6jt</a>	debaten denisedresserg sabinaberman deben continuar mañaneras campaña electoral	1	0	0
4715	“Toda persona merece tener a otra que la valore, que la respete, que la ame, y sobre todo que se lo demuestre.”	toda persona merece tener valore respete ame demuestre	0	1	0
1772	“Por semáforo rojo, clases de la UNAM en Edomex y CdMx seguirán siendo a distancia <a href="https://t.co/ywHVsqHFC">https://t.co/ywHVsqHFC</a> <a href="https://t.co/3SXbBOZXPz">https://t.co/3SXbBOZXPz</a> ”	semáforo rojo clases unam edomex cdmx seguirán siendo distancia	1	0	0
3448	Llorar es feo pero llorar de la bronca o del cansancio que te causa haber aguantado muchas cosas, eso es lo peor que existe.	llorar feo llorar bronca cansancio causa haber aguantado muchas cosas peor existe	0	1	0

Una vez que se limpiaron y estandarizaron, los textos del corpus fueron transformados a representaciones numéricas que, como se comentó previamente, permitan a la máquina codificar y decodificar de manera simultánea los datos textuales para su procesamiento y entendimiento (Sarkar, 2019). Para este caso se aplicaron, primero, la técnica de la TF-IDF que estima de manera automática la importancia de una palabra según la extensión del documento que la contiene (Vasiliev, 2020) y, posteriormente, la técnica de análisis de componentes principales (PCA, *Principal Components Analysis*), mediante la cual se reduce la alta dimensionalidad de un conjunto de datos al identificar sus variaciones respecto a la media (Beysolow II, 2018), permitiendo así centrar el análisis y procesamiento de los textos sólo en los atributos destacados por el PCA. Una vez implementadas las técnicas de transformación, al corpus se le añadieron las columnas de TF-IDF y de PCA (ver tabla 12).

Un análisis exploratorio sucinto (EDA) permite elaborar dos premisas generales respecto al corpus: la primera, que en los temas de tendencia (*trending topics*) en Twitter, predominaban —a la fecha en que se constituyó el corpus— el COVID-19, asuntos de política nacional e internacional y, en menor medida, el tema del amor y el gusto por la escritura, de lo que se puede inferir que los tweets catalogados como noticias tienen una mayor relevancia en la red, mas ello no implica que exista una mayor producción textual relacionada con este tipo de textos; segundo, tras el examen de frecuencias de palabras (ver figuras 23 y 24) se observa la recurrencia de palabras cuyo significado remite tanto a reflexiones como a relatos (tales como “vida”, “ser”, “tiempo”), lo que implica que exista una gran proximidad o semejanza léxica entre los textos catalogados como microficciones y los de reflexión.







Para comprender las diferencias, así como los atributos o características específicas de los microtextos que conforman este corpus se emplearon algunas de las aplicaciones de menor complejidad que se comentaron en la sección anterior, en concreto, las del etiquetado de partes del discurso (*POS tagging: Part of Speech*), la similitud entre textos, el análisis de dependencias sintácticas y el reconocimiento de entidades nombradas (*NER, Named-Entity Recognition*). Dada la gran cantidad de documentos con los que se cuenta para el análisis se seleccionaron observaciones de manera aleatoria para su examen particular o, en otras palabras, se escogieron al azar tres documentos etiquetados como “Noticia”, “Microficción” y “Frase/pensamiento”, respectivamente, los cuales fueron analizados mediante las aplicaciones de NLP mencionadas.

En lo que respecta a las noticias, el etiquetado de partes del discurso (*POS*) permite observar la preeminencia de formas verbales impersonales y de nombres propios, así como un uso mínimo de adjetivos, rasgos estilísticos acordes con los recursos retóricos de construcción de discursos factuales (Potter, 1998), mientras que el análisis de dependencias sintácticas (ver figura 25) evidencia una sencillez estructural en el documento, lo que implica que todo microtexto digital —o por lo menos todos los producidos en la plataforma de Twitter—, independientemente de si se trata de uno ficcional, informativo, motivacional, publicitario, etc., deben poseer las características de la brevedad y de la simplicidad sintáctica que se le atribuyen al microrrelato, para así garantizar una mayor eficacia comunicativa. Por último, el reconocimiento de entidades nombradas (*NER*) corrobora la presencia frecuente de vocablos que referencian objetos y sujetos reales (ver figura 26).

Figura 25. Análisis de dependencias sintácticas de tweets catalogados como noticias

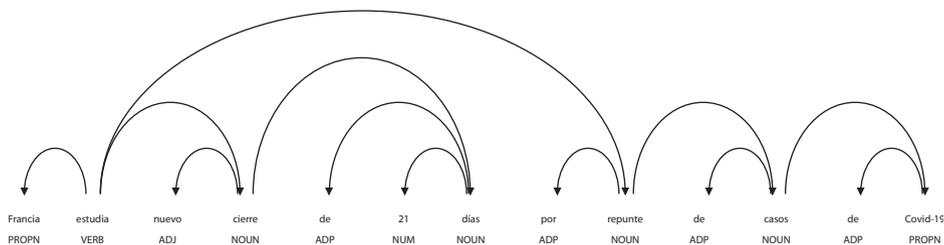
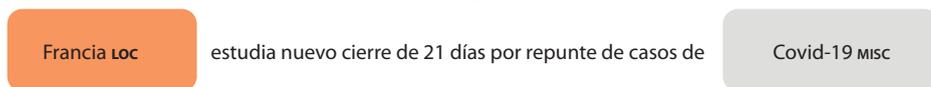


Figura 26. Reconocimiento de entidades nombradas (NER) en un tweet catalogado como noticia



Por su parte, en los tweets catalogados como frases o pensamientos se observa un mayor empleo de formas verbales personales, en un tiempo verbal presente, así como una mayor recurrencia de pronombres personales, rasgos de estilo que, posiblemente, se emplean para establecer un nexo empático con el lector y aludir al “aquí” y al “ahora”. Además, los atributos de brevedad y de simplicidad sintáctica parecieran no ser exclusivos de las microficciones, pues, de igual manera que con los microtextos comprendidos en la clase “Noticia”, los tweets que contienen frases o reflexiones tienden a la concisión y a una estructura sintáctica poco recargada (ver figura 27), aunque esto, como se señaló anteriormente, sean condiciones requeridas para la transmisión y difusión efectiva de mensajes a través de plataformas de microblogging (Russell y Klassen, 2019). No obstante, y como rasgo diferenciador entre los tweets informativos y los de “reflexión”, estos últimos no aluden a entidades nombradas.

En los tweets ficcionales, por otra parte, se observa el empleo de formas verbales personales conjugadas en tiempo pretérito, rasgo estilístico típico de los textos narrativos, además de un uso frecuente de pronombres personales; sin embargo, y a diferencia de los microtextos catalogados como noticias o frases/pensamientos, el análisis de dependencias sintácticas evidencia la presencia de oraciones subordinadas, usualmente con una función adjetiva, como un rasgo estructural recurrente en este tipo de microtextos (ver figura 28). Otro aspecto que se debe destacar es que los microrrelatos comprendidos en el corpus referencian de forma infrecuente a entidades nombradas. Como se puede observar de este somero examen, los atributos de la brevedad y de la simplicidad sintáctica, al ser comunes a las tres clases de microtextos analizadas en este estudio, no son criterios definitorios para clasificar a un tweet como microficcional.

Con base en lo anterior, el sistema de clasificación que más adelante se abordará se centra en los rasgos diferenciadores de la narratividad (Lagmanovich, 2006) y la convención estética, que conlleva a identificar como mi-

croficciones aquellos microtextos que no buscan representar de manera factual un evento —en claro contraste con los informativos—, que además brindan algún indicio para su lectura como microtextos con una clara y evidente intención literaria —establecen algún pacto tácito con el lector— y, por último, pueden aludir no sólo a un marco referencial “realista”, sino también a otros marcos referenciales —como por ejemplo, a mitos, leyendas o cultura popular— (Schmidt, 1997).

Recapitulando, dado que el análisis aislado de los diferentes tipos de microtextos compendiados en el corpus permitió ratificar que los atributos de la brevedad y la simplicidad sintáctica no son condiciones necesarias de distinción de los microrrelatos ficcionales respecto a otros microtextos digitales —por lo menos en lo que respecta a la plataforma de Twitter—, se ensaya a continuación el examen por similitudes léxicas y semánticas entre textos (Lane *et al.*, 2019) con el propósito de corroborar si los rasgos de la narratividad y de la convención estética son criterios suficientes (y evidentes) para construir el modelo en el que se basa el sistema de clasificación desarrollado en este estudio. Del análisis de similitud entre textos se espera definir, también, si los microtextos que conforman al corpus son representativos de tipos o clases textuales plenamente diferenciadas.

Para asegurar la reproductibilidad de este ensayo se retoman los microtextos que fueron examinados previamente mediante el POS, el análisis de dependencias sintácticas y el NER. Como este estudio se centra en la exploración de las características y atributos de los microrrelatos, específicamente de los tweets ficcionales, la comparación y búsqueda de similitudes entre microtextos se realiza cotejando los tweets catalogados como noticias y frases/pensamientos con los denominados como microficciones. Los microtextos comparados en este primer caso fueron “Francia estudia nuevo cierre de 21 días por repunte de casos de Covid-19”, perteneciente a la clase “Noticia”, con el microrrelato “Le ataron en medio del campo a unos tablores y le dejaron a su suerte, mecido por el viento. Lo peor es la soledad, pensaba el espantapájaros”, clasificado como microficción, de cuyo cotejo se obtuvo un porcentaje de 58% de similitud.

Figura 27. Análisis de dependencias sintácticas de tweets catalogados como frases o pensamientos

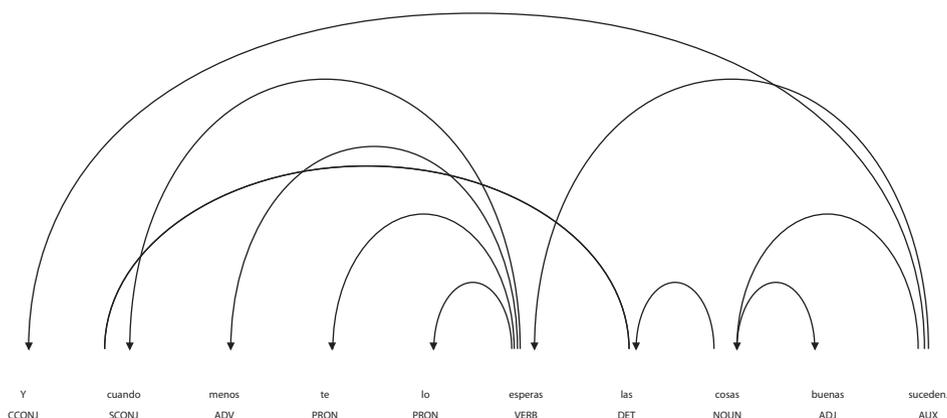
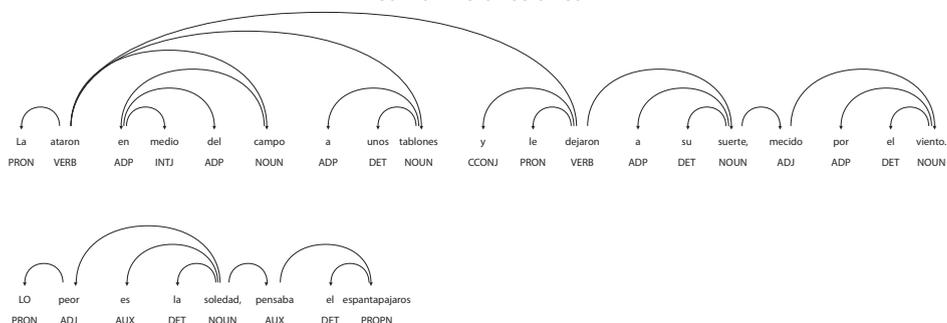


Figura 28. Análisis de dependencias sintácticas de tweets catalogados como microficciones



En lo que respecta a las similitudes entre los microtextos de frases o pensamientos con respecto a las microficciones, se comparó el tweet “Y cuando menos te lo esperas, las cosas buenas suceden”, con el microrrelato del ejemplo anterior y se obtuvo un porcentaje de similitud todavía menor que en el caso de las noticias, con 32%. De este examen se infiere que existe una mayor proximidad sintáctica y semántica entre los tweets informativos o de noticias y los ficcionales, dado que ambos tipos de texto poseen en común una tendencia narrativa o, en su defecto, *narrativizante*, esto es, una forma de articular el sentido del texto en función de una organización temporal definida —el texto remite a un evento en particular—, de una

estructura actancial relativamente estable —la situación o evento es propiciada o afecta a sujetos en particular— y de una fuente emisora tácita o explícita —el acontecimiento es relatado por un narrador, es decir, por una visión y una voz más o menos diferenciadas— (Barthes *et al.*, 1969; Escarpit, 1980).

En cuanto a las similitudes o diferencias entre los microtextos catalogados como frases o pensamientos respecto a las microficciones, la semejanza menor que se observa radica, posiblemente, en que una reflexión posee una estructura declaratoria, es decir, se trata de un enunciado o un conjunto de enunciados cuyo objetivo es afirmar o negar algo en particular; además, desde el punto de vista de la lógica, la retórica y la argumentación, los tweets de frases y reflexiones de este corpus se aproximan al tipo de silogismo denominado como entimema, en los que no existe una demostración como tal de la conclusión, al asumir la obviedad de las premisas que le anteceden (Perelman y Olbrechts-Tyteca, 1989), lo que significa que si bien puede existir una proximidad léxica entre las microficciones y los tweets de frases o pensamientos, éstos se distinguen sustancialmente en lo que concierne a sus estructuras sintácticas y sus rasgos semánticos.

Una vez que se han establecido estas distinciones entre las clases de microtextos que conforman el corpus, conviene ahora describir la aplicación con la que culmina este estudio y que, como se ha venido adelantando, consiste en un sistema de clasificación textual desarrollado mediante el análisis que se realizó anteriormente. Para la construcción de esta aplicación fue necesario “entrenar” un modelo, esto es, automatizar la detección de patrones en el corpus de microtextos (Albon, 2018); el proceso de entrenamiento, de acuerdo a los flujos básicos del aprendizaje máquina consistió en seccionar el conjunto de datos general para obtener dos subconjuntos, uno con 80% de las observaciones, que son los datos con los que se entrena el modelo, y el otro con 20% del total de microtextos del corpus, con el que se pone a prueba la eficacia del modelo tras su entrenamiento (Mathur, 2019).

El modelo de clasificación fue entrenado empleando Natural Language Toolkit (NLTK) (Bird, Klein y Loper, 2009) y, en mayor medida, spaCy (Honnibal, Montani, Van Landeghem y Boyd, 2020), dos de las librerías más populares de Python para el NLP. Con la finalidad de valorar su eficiencia

en las predicciones de casos diferentes a los del corpus, el modelo fue verificado con varios cientos de ejemplos de microtextos diferentes a los empleados para su entrenamiento; no obstante, el sistema clasificación de textos resultante de este estudio requiere múltiples mejoras, dado que en numerosas ocasiones no pudo clasificar varios de los microtextos con los que se probó<sup>10</sup> (ver figura 29). Para optimizar el modelo, y así mejorar las clasificaciones de textos nuevos o desconocidos para el sistema, se requiere o enriquecer el corpus con una variación y diversificación mayor de microtextos o también realizar un filtrado y un análisis más riguroso de los mismos o personalizar los parámetros de entrenamiento según los propósitos del modelado para así garantizar una precisión considerable en la categorización de los tipos textuales o, asimismo, implementar soluciones que combinen otras librerías diferentes de aprendizaje máquina y de procesamiento de lenguaje natural.

Figura 29. Captura de pantalla de la aplicación web de clasificación de textos.  
*La frase pertenece a Stephen Hawking*

En esta sección, puedes obtener información sobre la posible catalogación de un texto, según sea una microficción, una noticia o una reflexión

Escribe o copia y pega un fragmento de texto, y presiona Cmd y Enter/Ctrl y Enter para obtener la clasificación

Introduce un texto aquí:

El mayor enemigo del conocimiento no es la ignorancia, sino la ilusión del conocimiento

El texto no ha podido ser clasificado

Para concluir este capítulo es conveniente mencionar algunos de los usos prácticos que puede tener este sistema de clasificación en beneficio de artistas y creadores en general, así como para promotores, gestores y demás agentes involucrados en la producción cultural. Con las mejoras requeridas y con la adición adecuada de otras funcionalidades, la aplicación web aquí

<sup>10</sup> Se invita a los lectores y lectoras interesadas a probar la aplicación de clasificación de textos desarrollada a partir de este estudio accediendo al siguiente enlace: <https://share.streamlit.io/jeoa-1981/appmicrofiction/main/AppMicrorrelato.py>

presentada puede ser de utilidad no solamente para la clasificación de microtextos de ficción, de noticias o de reflexiones, sino también para otras clases textuales, tales como promociones o anuncios publicitarios, y cuyo análisis puede ser de provecho para la difusión del trabajo creativo, lo mismo que para eventos culturales.<sup>11</sup> De igual manera, se puede extender la aplicabilidad del sistema de clasificación textual para el análisis de sentimientos, esto es, para identificar las respuestas de los usuarios ante el contenido de un tweet en particular. Finalmente, si se complementa con un sistema de recomendación de filtrado colaborativo (Lope.ai, 2020) esta aplicación web puede ser de gran beneficio para escritores y escritoras emergentes.

---

<sup>11</sup> Cabe señalar que varias redes sociales, entre ellas Facebook, Instagram y el mismo Twitter, cuentan con sistemas de métricas y otras herramientas para impulsar el desarrollo de contenidos de sus usuarios; no obstante, sería pertinente investigar si este tipo de instrumentos son de conocimiento y de uso recurrente por parte de los diferentes agentes culturales y, en el caso de no ser así, fomentar su uso y entendimiento.



### **3. Visión computacional para el estudio de las artes plásticas**

#### **3.1. Iconografía, iconología y teoría de la imagen**

La visión es, con toda seguridad, una de las capacidades humanas esenciales, sino es que la más valorada, en tanto posibilita nuestra percepción del mundo. A través de la vista aprehendemos nuestro entorno, diferenciamos y comparamos los objetos que captan nuestra atención, percibimos los colores, las formas y las figuras en un paisaje, en una habitación o de un cuadro o escultura. Podría afirmarse sin mucha vacilación que nuestro sentido corporal más desarrollado es la vista y que su centralidad sensorial no radica solamente en que sea nuestro medio perceptivo principal, sino también en sus posibilidades expresivas y comunicativas, en el hecho de que seamos capaces de representar lo que vemos a través de mensajes visuales, de constructos que requieren ser comprendidos mediante la vista o, en palabras más sencillas, de imágenes.

Aunque la imagen ha sido siempre un medio expresivo y comunicativo de gran valor y utilidad ha tenido diferentes estimaciones a lo largo de la historia; al ser un constructo basado en nuestra capacidad sensorial básica, así como por la inmediatez con la que puede comunicar un mensaje, durante la Grecia Clásica la concepción filosófica generalizada de la imagen era negativa, dado que representaba solamente la exterioridad, el aspecto superficial y perecedero de las cosas y de las personas, en contraposición con las ideas, que remitían a lo racional, a lo permanente y sustancial (Zamora Águila, 2007). De manera semejante, durante la Edad Media la tradición judeocristiana condenaba las representaciones visuales que no remitieran a

motivos religiosos o santos, pues lo sensorial, lo físico, lo corporal alude a la exuberancia, al placer y, por consecuencia, al pecado (Debray, 2010).

Durante el Renacimiento —y los periodos históricos y culturales que le suceden— la imagen adquiere una significación diferente, en parte gracias a los debates entre las posturas empiristas y las racionalistas que derivaron en enfoques gnoseológicos de base sensualista o de orientación idealista, respectivamente. (Zamora Águila, 2007). A través de las imágenes se pueden transmitir con facilidad toda clase de ideas y valores, ya sean políticas, religiosas o filosóficas, normas de etiqueta o de conducta, etc.; por tal razón, los productores de imágenes, conscientes de los diversos usos de éstas, adquirieron durante estas épocas una mayor presencia en la vida pública, pues a partir de su trabajo eran capaces de influir en la percepción e interpretación del pueblo, en favor o perjuicio de un régimen político o religioso, interviniendo así en la preservación o en la destrucción del *statu quo* (Debray, 2010).

Hoy en día, y gracias a las nuevas tecnologías de la información y la comunicación, la imagen mantiene su carácter preeminente por sobre otras formas o medios expresivos y comunicativos, además de que la velocidad con la que se producen y se propagan las imágenes se ha incrementado exponencialmente con respecto a las épocas antecedentes (Jullier, 2004). La imagen se ha tornado en algo tan cotidiano, tan inmediato y tan fugaz que su conservación —su captación y rememoración por parte del espectador— depende en gran medida de que sea capaz de destacarse de entre todas las imágenes semejantes o parecidas, en una especie de darwinismo icónico (Gubern, 1996); además, el acceso cada vez más generalizado a dispositivos tales como cámaras fotográficas y de video, integradas en un mismo artefacto, el teléfono móvil, o a computadoras personales y *software* de creación y edición de imágenes intensifica lo que Gubern sugería como una competencia feroz de las imágenes por nuestra atención.

Tras este muy sucinto repaso histórico en torno a las concepciones y usos de la imagen, es pertinente esclarecer a lo que este término alude, esto es, definir qué es una imagen, qué tipos de imágenes existen, además de los enfoques y propuestas analíticas a las que se recurre para su entendimiento. En su acepción más cotidiana el vocablo de imagen remite a las referencias o representaciones de la realidad construidas a partir de elementos visuales,

esto es, que requieren de la vista para ser comprendidos tanto en su particularidad como en su generalidad (Zunzunegui Díez, 2010); no obstante, existen imágenes cuya concepción no requiere de la presencia efectiva del objeto, como son las imágenes mentales, que aluden por lo general a recuerdos, fantasías e incluso a sensaciones en particular (Zamora Águila, 2007), como, por ejemplo, un aroma determinado que puede evocar el sabor de un alimento, un episodio de la infancia, una situación posible o deseada —o no deseada— en un futuro.

Las imágenes mentales se catalogan de acuerdo con la capacidad sensorial a la que referencian, existiendo así las visuales —que evocan la percepción visual de un objeto, de un acontecimiento o de un entorno—, las auditivas, las gustativas, las olfativas, las táctiles y las cinestésicas (Zunzunegui Díez, 2010), estas últimas son las que simulan o reproducen la sensación corporal del movimiento y son de las más populares en el campo de la realidad virtual (Jullier, 2004). La literatura y las artes en general, lo mismo que los medios de comunicación, se sirven con frecuencia de las imágenes mentales para construir visiones poéticas, para evocar situaciones o, incluso, para construir marcos referenciales inéditos o mundos ficcionales (Campra, 2008). La imagen, en su sentido más general, no se puede disociar de las operaciones ni del funcionamiento de la psique humana.

Sin embargo, usualmente la imagen se relaciona con lo visto, con lo que es percibido de manera actual y efectiva mediante el sentido de la vista; desde esta perspectiva, el concepto de imagen pareciera ser elaborado generalmente a partir de criterios materialistas, esto es, en términos de su exterioridad, ya sea en función del proceso, técnicas o materiales por medio de los cuales se construye o se elabora, como es el caso del dibujo, el garabato, el grabado, la pintura, la fotografía, el cine y el video —imágenes o conjuntos de imágenes en las que interviene en un grado mayor una tecnología— (Zunzunegui Díez, 2010), o bien, según su soporte, que puede ser analógico o digital (Jullier, 2004) o, también, en la valoración o función que se le otorga, independientemente de su modo o medio de producción (Joly, 2012), en cuyo caso la imagen se clasifica en artística, publicitaria, científica, educativa, etcétera.

Sea cual sea el origen de la imagen es evidente su gran relevancia, tanto si se le concibe como un constructo cultural que media entre lo objetivo y

lo subjetivo, entre la percepción y la interpretación (Zamora Águila, 2007) o, de manera complementaria y consecuente, como una forma de producción de sentido, como un artefacto semiótico cuyo potencial cognitivo consiste precisamente en brindarnos una representación del mundo empírico (Joly, 2012). Al vivir inmersos en un mundo repleto de imágenes de toda clase que se producen, reproducen y proliferan con una velocidad excepcional (Gubern, 1996), su exploración, análisis y entendimiento se presenta como una tarea que, aunque compleja, es necesaria no sólo por su valor y sus cualidades como artefactos perceptivos y cognitivos (Zunzunegui Díez, 2010), sino también porque su examen pudiera ayudar a comprender en alguna medida las complejas dinámicas culturales contemporáneas.

Los enfoques teóricos en torno a la imagen parten de planteamientos diversos, aunque no necesariamente contrapuestos. El estudio de las representaciones o constructos visuales fue, durante mucho tiempo, jurisdicción de la filosofía, concretamente de la gnoseología y de la estética (Zamora Águila, 2007); en el primer caso, la reflexión gnoseológica se centraba, como se sugirió al inicio de este capítulo, en las dimensiones referenciales de la imagen, es decir, en si la imagen referenciaba o no a un objeto y, por lo tanto, en si remitía a un conocimiento sobre éste o si conducía a la falsedad y al error (Debray, 2010); por su parte, la estética abordaba las cualidades sensoriales de la imagen —en particular, de la artística—, así como los efectos que ésta tenía en nuestra sensibilidad —si una imagen nos conmueve, nos entristece, nos alegra, etc.—; cabe señalar que, históricamente, las artes y las imágenes han estado estrechamente asociadas, tal vez porque, desde Platón, el trabajo artístico se concebía como una actividad orientada a la creación de imágenes, y estas últimas se consideraban principalmente como representaciones de lo captado por los sentidos (Tatarkiewicz, 2002).

Desde finales del siglo XIX el estudio de la imagen se ha ampliado, admitiendo así planteamientos de la psicología, la lingüística, la teoría de la comunicación y de la semiótica, siendo este último el campo en el que se han integrado las propuestas de múltiples disciplinas. Sucintamente, la semiótica es una ciencia generalista que investiga los diferentes sistemas sígnicos, sus medios y modos de producción, los intercambios que se dan entre los sujetos del acto comunicativo, así como los códigos por medio de los cuales se comprenden los signos (Eco, 2000); la imagen, en este caso, es

un tipo de signo, es decir, una unidad conformada por un aspecto material, o dimensión significativa, que remite a un referente exterior, y un contenido, o significado, que se refiere a la ideación por parte del intérprete —del sujeto que recibe el mensaje—.

Según la relación que el significante —esto es, la materialidad del signo visual— establezca con su referente, las imágenes se clasifican en iconos, índices y símbolos. Un icono sería la clase de signo visual que es análogo a su referente o, en otras palabras, que representa lo más fidedignamente posible a su objeto —como, por ejemplo, la pintura y la escultura de tradición figurativa, que buscan imitar a la naturaleza y a la figura humana (Tartakiewicz, 2002)—; por su parte, se considera como un índice a un tipo de signos que aluden de manera causal a su referente, esto es, que expresan el efecto de algún suceso o evento —como el pavimento mojado, que indica que ha llovido recientemente—; en lo que respecta al símbolo, se trata de los signos que, aunque no sean semejantes a su referente, guardan con éste una relación por convención o por connotación —por ejemplo, la manzana, desde la visión católica, es un símbolo de la desgracia del género humano— (Joly, 2012).

Si bien esta es una de las tipologías más tradicionales en la semiótica, su vigencia radica en que permite identificar las características predominantes en un mensaje visual; como bien señala Joly, no existen signos puros, antes bien, cada uno posee un cierto carácter o nivel de iconicidad, lo mismo que de indicidad —o indexicalidad— y de simbolismo, como se observará más adelante en este capítulo. Con base en los aspectos resaltados por esta clasificación, el análisis de una imagen se centra en la identificación y estudio de sus elementos constitutivos, que varían según el modo y medio de producción visual; en el caso de la pintura, por ejemplo, algunos de sus elementos considerados como esenciales son la forma, el color, la composición y la textura, mientras que en la fotografía y en el cine son el plano, la secuencialidad, el encuadre y el montaje los componentes que a menudo reciben mayor atención en el análisis (Joly, 2012; Zunzunegui Díez, 2010).

Por su enfoque abarcador la semiótica visual tiene por objeto de estudio a toda clase de imágenes, indistintamente de si se trata de imágenes artísticas —como pinturas, ilustraciones, grabados—, lúdicas o de esparcimiento

—como los memes o las tiras cómicas—, publicitarias —anuncios, carteles—, etc., no obstante, y por la propensión interdisciplinaria de la semiótica general, el análisis de la imagen se apoya en herramientas de disciplinas específicas al tipo de constructo visual o imagen que se desea examinar al mismo tiempo que las integra. En lo que concierne a la imagen artística, la semiótica visual se apoya en la estética, lo mismo que en la teoría, la crítica y la historia del arte; precisamente, uno de los aportes más provechosos para el estudio semiótico de las representaciones visuales que se catalogan como artísticas es el del método iconográfico-iconológico del historiador del arte Erwin Panofsky.

La iconografía, según la describe el mismo Panofsky, es una vertiente de la historia del arte enfocada en el examen del “asunto” o significación de la obra artística; el autor distingue tres niveles de significación en el arte, el primero referente a la captación “ingenua” o natural de los elementos visuales básicos y de sus relaciones al interior de la obra —formas, objetos, acontecimientos—, designada como *significación primaria* o natural, que a su vez comprende una *significación fáctica* o factual y una *expresiva*; la segunda, nombrada como *significación convencional*, se refiere a la identificación de connotaciones, esto es, de motivos asociados con narrativas míticas o alegóricas; por último, el tercer tipo es el de la *significación intrínseca* o de *contenido*, que consiste en la identificación y enunciación del “espíritu de la época”, de las ideas, valores y normas predominantes del periodo y de la nación de la que se deriva la obra de arte (Panofsky, 2011, pp. 47-51).

Cada uno de estos niveles de significación conlleva a un tipo de descripción de la obra de arte; partiendo de su estrato más elemental, la primera descripción, que es más bien una enumeración de motivos factuales y expresivos, es pre-iconográfica, y a ésta le sucede una iconográfica como tal, que consiste en la referenciación de alegorías, mitos e historias, y se concluye con la iconológica, que es la descripción que articula todos los motivos o asuntos presentes en la obra artística en una interpretación que posibilite su comprensión contextual, esto es, su historicidad —como ejemplar de un *estilo* o movimiento artístico, lo mismo que como resultado de transformaciones y evoluciones culturales—. Es pertinente mencionar que el método propuesto por Panofsky no ha estado exento de críticas, no obstante, su utilidad para el análisis de la imagen en general, como de la imagen artísti-

ca en particular, reside en la centralidad que el significado adquiere para la lectura de la obra de arte (Schirato y Webb, 2004).

Esta breve revisión de las concepciones en torno a la imagen, así como de los planteamientos teórico-metodológicos para su abordaje, conduce ahora a detallar el estudio que aquí se desarrolla. El propósito de esta investigación, a diferencia de las presentadas en los capítulos previos, no es analizar un fenómeno cultural con la asistencia de herramientas tecnológicas y de métodos y técnicas de la ciencia y de la analítica de datos, aunque se mantiene la aproximación reflexiva de la analítica cultural (Manovich, 2020); antes bien, el doble objetivo de este estudio es el de ilustrar los usos y las aplicaciones generales de la disciplina científica denominada visión computacional (*Computer Vision*) y, de ahí, explorar sus posibles empleos para el análisis de imágenes artísticas, concretamente de la plástica, que incluye a la pintura, la escultura y el grabado, entre otras manifestaciones artísticas que parten de la manipulación de materiales para la creación de obras (Tatarkiewicz, 2002).

Un aspecto de carácter terminológico que conviene aclarar es el del empleo de la noción de artes plásticas en lugar de artes visuales. Aunque la mayoría de los estudios contemporáneos en torno al arte se decantan por el término de artes visuales, que es más comprehensivo que el de artes plásticas (Schirato y Webb, 2004), para este estudio se ha optado por este último por una razón fundamental: el espectro de prácticas artísticas incluidas en las artes visuales comprenden a menudo relaciones interartísticas o intermediales, esto es, conexiones e interacciones entre sistemas artísticos diferentes —como, por ejemplo, la escultura con el teatro o la literatura con la pintura— (Herzogenrath, 2012) y que, desafortunadamente, superan las posibilidades actuales de este estudio, por lo que se optó por una conceptualización que, aunque esquemática, es de utilidad para los fines de esta investigación.

Así pues, este estudio parte de una interrogante concreta, que es la de si las técnicas de procesamiento de imágenes pertenecientes a la visión computacional posibilitan —y facilitan— un análisis más detallado de imágenes de obras artísticas, lo que implica preguntarse también si a través de la visión computacional es posible identificar los elementos esenciales de una obra de arte de la plástica. Para responder a estos cuestionamientos se desarrolla

una aplicación web de visión computacional que será probada con imágenes artísticas con licencia *Creative Commons*<sup>12</sup> con la intención de identificar de manera automatizada los elementos visuales esenciales de la obra —forma, color, composición y textura, u otros, según se requiera (Joly, 2012; Zunzunegui Díez, 2010)— y, a partir de ello, ensayar un aproximación a ésta mediante el método iconográfico-iconológico de Panofsky.

### 3.2. Visión computacional: usos académicos e industriales

La visión computacional, también llamada visión máquina o visión artificial, es una disciplina que hoy en día ha adquirido un lugar destacado no sólo en numerosos sectores industriales, lo mismo que en la docencia y la investigación científica, sino también en la vida cotidiana. Al igual que el NLP, cuyas variadas aplicaciones —tan insertas ya en la cotidianidad— se dan por sentado, la visión computacional se ha integrado con extrema facilidad a las actividades humanas más ordinarias y habituales a tal grado que, independientemente de la complejidad de sus múltiples implementaciones o de su novedad su presencia pasa desapercibida en una sociedad familiarizada con las imágenes y, en ocasiones, incluso abrumada e indiferente ante su sobreabundancia o exceso.

Algunas de las aplicaciones más “sencillas”<sup>13</sup> de la visión computacional, y con las que con mayor frecuencia se llega a tener contacto al navegar en internet, son, por ejemplo, los captcha (Completely Automated Public Turing test to tell Computers and Humans Apart), pruebas digitales que se emplean para permitir el acceso a una página —o para descarga de datos— en las que, a manera de desafío, se le pide al usuario identificar objetos específicos en imágenes, como una forma de “verificar” que se trata de un ser humano; los sistemas de reconocimiento y de codificación y de decodificación de caracteres y de símbolos, como los códigos de barra, tan comunes

<sup>12</sup> Es una de las licencias digitales más permisivas, pues admite un reuso tanto académico como comercial (<https://creativecommons.org/>).

<sup>13</sup> En el sentido de que en la actualidad se pueden desarrollar incluso con conocimientos básicos de programación y código, o mediante el uso de algún *software* en particular, además de contar con numerosos tutoriales para reproducir su confección.

en los supermercados, o los QR (*Quick Response*), cada vez más usuales para acceder a información en tiempo real sobre un producto —por ejemplo, el valor nutricional de un alimento o las características de un electrodoméstico— o un servicio —trámites gubernamentales o compras en línea— son aplicaciones de la visión computacional cuyo uso es muy cotidiano.

Otras implementaciones un tanto más complejas de la visión computacional van desde los motores de búsqueda visual, en los que el usuario introduce una imagen de referencia y el sistema devuelve una cifra determinada de coincidencias visuales que concuerden con las características más relevantes de la imagen ingresada (gama cromática, luminosidad, objetos en primer plano, etc.), el reconocimiento facial, cada vez más utilizado en algunos modelos de teléfonos inteligentes o *smartphones*, hasta el reconocimiento de gestos y movimientos, que es una de las bases de los gráficos generados por computadora (*CGI, Computer Generated Imagery*), muy recurrente en la industria del cine y en la de los videojuegos. Ahora bien, algunos de los usos más sofisticados y complejos de la visión computacional se dan, por ejemplo, en la medicina, con el diagnóstico asistido en reconocimiento de imágenes,<sup>14</sup> o en la industria del transporte y la milicia, con los vehículos autodirigidos, desde coches y camiones hasta drones y aviones, mientras que en el entretenimiento, las aplicaciones de realidad aumentada (*Augmented Reality*) y de realidad virtual (*Virtual Reality*) adquieren cada vez una mayor popularidad.<sup>15</sup>

Como puede observarse de lo anterior, la visión computacional es un terreno de gran atractivo para el sector público o gubernamental, lo mismo que para el empresarial e industrial, y el académico, sin olvidar los beneficios que puede reportar a la sociedad en general, aunque esto no debe entenderse como una alabanza acrítica de una tecnología, o conjunto de tecnologías, que pueden derivar también en regímenes de vigilancia permanente, a manera de un panóptico o de un Gran Hermano que opera con inteligencia artificial; precisamente para prevenir este tipo de problemáticas —o, en su defecto, resolverlas— se requiere comprender desde una óptica integral, esto es, ética, estética, social y ecológica, las diversas tecnologías con las que

<sup>14</sup> Ver Al-antari, Hua, Bang y Lee, (2020).

<sup>15</sup> Recuérdese el fenómeno de Pokémon GO, que desde el 2016 y hasta el 2019 se mantuvo como uno de los juegos de realidad aumentada más populares de todo el mundo.

funciona nuestro mundo, y con vistas a la equidad y a la mejora democrática de la calidad de vida de los seres humanos (Schwab, 2018).

Con miras a entender la visión computacional desde la perspectiva de la analítica cultural, esto es, considerando su valor para el estudio masivo de toda clase de imágenes digitales, lo mismo que para proyectar sus usos y aplicaciones en beneficio de las actividades artísticas y culturales en general, conviene conocer, aunque sea de manera sucinta, en qué consiste este ámbito interdisciplinario, de qué técnicas se sirve, lo mismo que sus alcances y limitaciones. En la década de 1960 en las universidades se daban los primeros pasos para el desarrollo de la visión computacional, cuyo propósito era lograr que las computadoras emularan la visión humana (Ballard, 2018; Maria Farinella, Battiato y Cipolla, 2013); a partir de técnicas todavía rudimentarias para el procesamiento de imágenes, y en conjunto con los avances posteriores en inteligencia artificial y aprendizaje máquina (Mathur, 2019), la visión computacional adquirió paulatinamente el perfil de una disciplina permanentemente innovadora (Maria Farinella *et al.*, 2013).

La visión computacional se basa en la identificación y análisis automático de los atributos visuales existentes en una imagen, esto es, las propiedades mediante las cuales se pueden distinguir objetos reales o abstractos en un mensaje visual (Dawson-Howe, 2014); según Weinmann (2013), los atributos visuales pertinentes o valiosos para la visión computacional debieran cumplir con cinco condiciones específicas: primero, deben ser *distintivos*, esto es, que difieran significativamente respecto de otros atributos presentes en una imagen; segundo, deben ser *invariantes*, lo que implica que no tengan mucho “ruido” visual —variaciones aleatorias de luminosidad o colores que afectan la nitidez de la imagen— (Gollapudi, 2019); tercero, que sean *eficientes*, o lo que es lo mismo, que no requieran un enorme esfuerzo computacional para su procesamiento; cuarto, que sean comparables con respecto a otros atributos presentes en la misma o en otras imágenes; y, por último, que sean relevantes, esto es, que conduzcan a algún hallazgo o conocimiento respecto al objeto visual (Weinmann, 2013, p. 7).

Un descubrimiento afortunado para este estudio radica en el hecho de que los atributos visuales en los que se centran las técnicas de visión computacional coinciden en su mayoría con los elementos visuales en los que

se basa el análisis de la imagen en general (Joly, 2012; Zunzunegui Díez, 2010), como lo son el color o la intensidad cromática, la forma y la textura de los objetos presentes en la representación visual, que se obtienen o se infieren a través de la detección de bordes y contornos (Gollapudi, 2019) —como se verá en la siguiente sección de este capítulo—, el entorno, entendido como la visión de conjunto de las características previamente comentadas, y el movimiento, efectivo —como en un video— o sugerido —como en una fotografía—; estos dos últimos aspectos se consideran atributos espacio-temporales generalizadores (Weinmann, 2013).

Las técnicas mediante las cuales opera la visión computacional se pueden clasificar en, primero, operaciones de filtrado, consistentes en la reducción del ruido, la mejora de la nitidez de un objeto o de la imagen en su totalidad, el “suavizado” y el “agudizamiento” o “afilado” de texturas o bordes pixelados, entre otras más (Howse, Joshi y Beyeler, 2016); segundo, la visión computacional también se sirve del seguimiento o rastreo de características o atributos visuales, o de objetos usualmente en videos o en imágenes de organización secuencial, como cómics o fotonovelas (Dawson-Howe, 2014); tercero, las operaciones de detección en la visión computacional van desde la localización de atributos simples, como el color o la forma, de objetos en una imagen, hasta la identificación de conjuntos de características o atributos visuales, por ejemplo, en la distinción de partes del rostro<sup>16</sup> o del cuerpo humano o de algún objeto en particular (Gollapudi, 2019) —como las placas de un automóvil o los signos de una señal de tráfico—; cuarto, la extracción de atributos visuales, que consiste en la recolección de características esenciales en la imagen para su procesado y uso posterior en una aplicación (Ballard, 2018).

Sin intención de simplificar, de la combinación de las técnicas y operaciones previamente comentadas se originan las diversas aplicaciones con las que la visión computacional se hace presente en casi todos los sectores productivos de la sociedad actual; a título de inventario, y también para complementar el inicio de esta sección, es pertinente compendiar, aunque sea de manera sintética, los ámbitos en los que la visión computacional ha tenido cabida, ya sea para innovación o mejora, o bien como detonante de

---

<sup>16</sup> Para un ejemplo práctico de estas operaciones de visión computacional ver Wattamwar, 2021.

un nuevo campo productivo. Como se sugirió antes, varios avances medicinales se basan precisamente en la aplicación de técnicas de procesamiento de imágenes para la detección de anomalías que pudieran afectar la salud del paciente; por otra parte, el reconocimiento facial es ya una alternativa de seguridad implementada por algunos bancos<sup>17</sup> (Mathur, 2019).

Sin duda, la industria manufactera es uno de los sectores más beneficiados por la visión computacional, mediante la cual ha sido posible desarrollar robots capaces de reconocer la materia prima para la elaboración de un producto determinado, así como de seguir su confección en cada fase de producción; aunado a ello, también hay aplicaciones industriales para el monitoreo general del producto y para la detección de fallos o defectos en particular (Howse *et al.*, 2016); una implementación parecida, y que se mencionó con anterioridad, se da en el sector de transportes con el uso cada vez más sistemático de sistemas de entrega automatizados por medio de drones que pueden reconocer objetos diversos o de sistemas de autodirección asistida, como es el caso de algunos automóviles (Gollapudi, 2019). Incluso en el ámbito deportivo es frecuente el uso de técnicas para el seguimiento de objetos (*object tracking*).

En la agricultura, por su parte, se han diseñado sistemas de estimación para la calidad de la cosecha, lo mismo que para atender las necesidades de los plantíos (Mathur, 2019). La visión computacional también ha sido de utilidad en el manejo y/o reciclaje de desechos, empleando técnicas de identificación y distinción de objetos o componentes que puedan ser reutilizados, lo mismo que en la valoración y/o pronóstico de los efectos de desastres naturales en entornos urbanos (Gollapudi, 2019).<sup>18</sup> En lo que respecta a las artes, la cultura y las industrias creativas en general, dada su intrínseca tendencia a la innovación, las diferentes operaciones de procesamiento automatizado de imágenes, las aplicaciones de la realidad aumentada y de la realidad virtual son habituales en la experimentación artística y en la generación de piezas intermediales (Elleström, 2010; Herzogenrath, 2012), lo mismo que para la creación y modificación de contenidos en los medios (Jenkins, Beyeler y Green, 2015).

<sup>17</sup> Un ejemplo es el del registro de reconocimiento facial empleado por BBVA para identificar a sus cuentahabientes (<https://www.bbva.mx/personas/faq/landings/reconocimiento-facial.html>).

<sup>18</sup> Ver Valente, Silva, Caldeira, Soares, y Gaspar (2019); y Yogameena y Nagananthini (2017).

En lo que compete a este estudio, como se ha indicado en anteriores ocasiones, su propósito es comprobar si las técnicas de procesamiento de la visión computacional son de utilidad para la identificación de los elementos esenciales en una obra de arte de la plástica, para así facilitar y agilizar el análisis de la imagen artística; para ello, y al igual que en los capítulos previos, se ha desarrollado una aplicación web para ilustrar los usos de la visión computacional con piezas selectas de las artes plásticas. Las operaciones comprendidas en la aplicación corresponden a la identificación de objetos con el modelo MobileNet SSD, que se basa en una red neuronal de aprendizaje profundo (*Deep Learning*) diseñada para la detección de objetos mediante la localización de atributos diferenciadores y su separación con base a “cajas” contenedoras predefinidas (Liu *et al.*, 2016).

Se incluye también en esta aplicación una técnica de detección de bordes basada en el algoritmo de Canny, desarrollado por John F. Canny en 1986. El funcionamiento de este algoritmo consiste en, primero, llevar a cabo un filtrado y reducción del ruido o variaciones aleatorias de la luminosidad y del color en la imagen; después, se localizan los gradientes de intensidad, esto es, las zonas de mayor concentración o intensidad luminosa o cromática y, finalmente, se remueven y se suprimen los píxeles innecesarios en la imagen; aunado a lo anterior, se incorpora además una operación para la detección de contornos, que se definen como las líneas en las que convergen todos los puntos de similar intensidad que conforman la figura de un objeto (Gollapudi, 2019). En el estudio de imágenes artísticas, la detección de contornos podría ser de utilidad para analizar las formas y las texturas predominantes en una obra o en un conjunto de piezas artísticas.

Es importante subrayar que tanto la identificación de objetos como la detección de bordes y contornos funcionan con base en modelos pre-entrenados o en algoritmos específicos, por lo que su precisión para el análisis de imágenes de obras artísticas no es en absoluto el deseado; no obstante, se insiste en que, en este caso en particular, el estudio tiene una finalidad específicamente ilustrativa. Por último, la aplicación cuenta también con un modelo de clasificación de imágenes desarrollado con la librería Tensorflow (Abadi *et al.*, 2016), el cual fue entrenado con un corpus de cerca de 9 000 imágenes correspondientes a dibujos, grabados, iconografías, pinturas y

esculturas.<sup>19</sup> Sin embargo, el sistema de clasificación resultante es todavía inexacto y requiere ser re-entrenando con ejemplos más diversificados o, en su defecto, se necesita ajustar los parámetros de entrenamiento según las características de las imágenes a clasificar.

### **3.3. Análisis de imágenes artísticas asistido por visión computacional**

Para demostrar el potencial de la visión computacional para el análisis de obras artísticas y, de manera conjunta, el funcionamiento de la aplicación desarrollada en este estudio, se emplean, respectivamente, ejemplos pictóricos de tradición figurativa y abstracta tanto para la identificación de objetos como para la detección de bordes y contornos; en lo que respecta a la clasificación de imágenes, el modelo se prueba con un ejemplo perteneciente a cada una de las categorías con las que fue entrenado (dibujos, grabados, iconografías, pinturas y esculturas). Cabe agregar que la técnica de identificación de objetos sería de gran utilidad para localizar motivos artísticos que permitan ubicar el estilo histórico al que pertenece —o podría pertenecer— la obra; por su parte, la detección de bordes y contornos puede brindar información de relevancia para definir ciertas características estilísticas de la obra —la predominancia de luces y sombras, el empleo de los colores, si son formas orgánicas o inorgánicas las que destacan en la pieza, etc.—. La clasificación automatizada de imágenes, por otra parte, sería —o es, más bien— de gran provecho para la catalogación y peritaje de obras en museos y galerías de arte.

La función para identificar objetos fue probada con variadas piezas pictóricas, algunas pertenecientes al renacimiento y otras a los movimientos artísticos del surrealismo, del cubismo y del expresionismo abstracto, para así explorar las diferencias resultantes entre el examen de imágenes que representan a sus objetos y figuras con una mayor nitidez y claridad, y las que, por otra parte, abstraen las propiedades que se consideran inesenciales o, en su defecto, que simbolizan conceptos o ideas. Como era de esperarse,

<sup>19</sup> El modelo fue provisto, de manera libre y abierta, por el usuario "Danil", en el sitio web Kaggle: <https://www.kaggle.com/thedownhill/art-images-drawings-painting-sculpture-engraving>

y dado que el modelo con que se realiza esta operación no fue entrenado con imágenes de obras artísticas, los resultados de la identificación de objetos fueron inexactos, pudiendo distinguir mayormente las siluetas humanas representadas en primer plano o con atributos visuales más detallados (ver figura 30) y, en otros casos, se confundían cosas con personas o con el fondo mismo (ver figura 31) o, incluso, no se distinguía ningún objeto (ver figura 32).

En cuanto a las piezas de movimientos tendientes a la abstracción, en una primera prueba se obtuvo una confusión respecto a los objetos existentes en la imagen (ver figura 33), mientras que en los siguientes intentos, el modelo no pudo identificar ningún objeto (ver figuras 34 y 35); evidentemente, el nivel de iconicidad en una obra artística determina por completo la distinción de objetos mediante técnicas de visión computacional, y pese a que en este caso la operación de identificación automatizada de objetos no produjo resultados muy certeros, es pertinente reiterar que el entrenamiento de un modelo específico para la detección de motivos artísticos —basado precisamente en técnicas de identificación de objetos— sería de gran utilidad en el análisis de imágenes artísticas, principalmente como auxiliar en descripciones pre-iconográficas de obras artísticas (Panofsky, 2011), esto es, en el inventariado de detalles icónicos que aludan o que sean referentes a objetos míticos, legendarios o históricos. No obstante, para desarrollar y entrenar un modelo de ese tipo sería necesario, primero, una extensa investigación histórica en torno a los motivos artísticos empleados en diferentes épocas y estilos (Schirato y Webb, 2004).

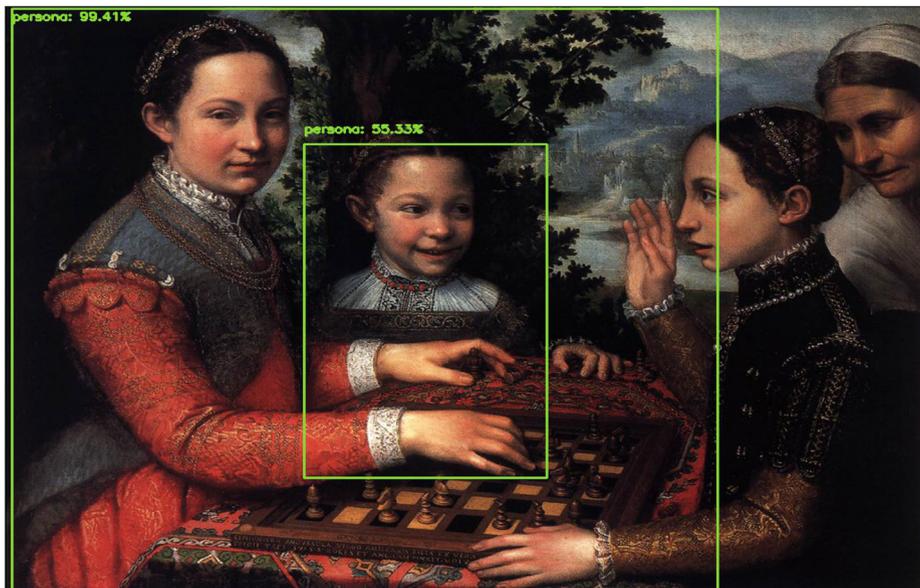


Figura 30. Sofonisba Anguissola, *Las hermanas de la artista*, 1555. En esta pieza sólo se distingue al personaje que ocupa un plano más relevante o, en su defecto, al que tiene un rostro más definido



Figura 31. Sandro Botticelli, *La calumnia de Apeles*, 1495. En esta pieza sólo se distingue un personaje humano, pero el sistema de identificación cataloga a los del centro con la etiqueta "silla".



Figura 32. Remedios Varo, *Mujer saliendo del psicoanalista*, 1960. En este caso el modelo no pudo identificar ni siquiera figuras de personas.



Figura 33. Pablo Picasso, *El marino*, 1943. El modelo de identificación confunde la figura humana con una silla, posiblemente por su postura.



Figura 34. Jackson Pollock, *Head*, 1938-1941. *En lo que respecta a esta pieza, no se pudo identificar ninguna clase de objeto.*



Figura 35. Cy Twombly, *Untitled (Bacchus)*, 2005. *Por el nivel de abstracción de esta pieza, el modelo no pudo distinguir objeto alguno.*

La comprobación de la operación de detección de bordes y contornos se realizó empleando también ejemplos pictóricos de mayor a menor nivel de iconicidad, es decir, de cuadros figurativos a abstractos. A diferencia de la técnica de identificación de objetos, con la detección de bordes y contornos se obtuvieron aproximaciones un tanto más acertadas a las piezas procesadas. En el primer caso (ver figuras 36 y 37), la localización de bordes y contornos permitió corroborar el uso del claroscuro en la obra, resultado del contraste entre el predominio de sombras y la concentración de los valores cromáticos y lumínicos de manera parcial en algunos puntos del espacio pictórico de la obra (Zunzunegui Díez, 2010). En esta primera prueba, al tratarse de una pieza de Caravaggio, la antítesis entre luces y sombras tiene una intención claramente dramática, para enfatizar el padecimiento o la vehemencia del episodio mítico o histórico reproducido.

Para la siguiente comprobación se empleó otra obra de Remedios Varo, de cuyo procesamiento se observa una mayor uniformidad de valores cromáticos y lumínicos (ver figuras 38 y 39), tanto en lo que respecta a la composición de la figura o personaje central en el cuadro como en la representación de los objetos que conforman la composición en general (Joly, 2012); mientras que en la pieza de Caravaggio la detección de bordes y contornos impele a inferir que en obras de tipo figurativo es usual la concentración cromática o la densidad lumínica en puntos específicos de la imagen, en el caso del cuadro de Varo podría pensarse que la distribución más o menos equitativa de bordes y contornos es resultado de una orientación pictórica hacia el aspecto simbólico antes que al figurativo (Schirato y Webb, 2004).

El último ejemplo corresponde a una pieza de la artista Susan Rothenberg, cuyos trabajos tienden a la búsqueda de una figuración no constreñida por referentes realistas, lo mismo que a una metaforización de los objetos por intermedio de la primacía de las formas; en este caso, la detección de bordes y contornos permitió verificar una mayor concentración cromática en secciones específicas del cuadro, así como una nula presencia de contrastes entre luces y sombras.



Figura 36. Detección de bordes de la pintura *David con la cabeza de Goliat* (1609-1610), de Michelangelo Merisi da Caravaggio

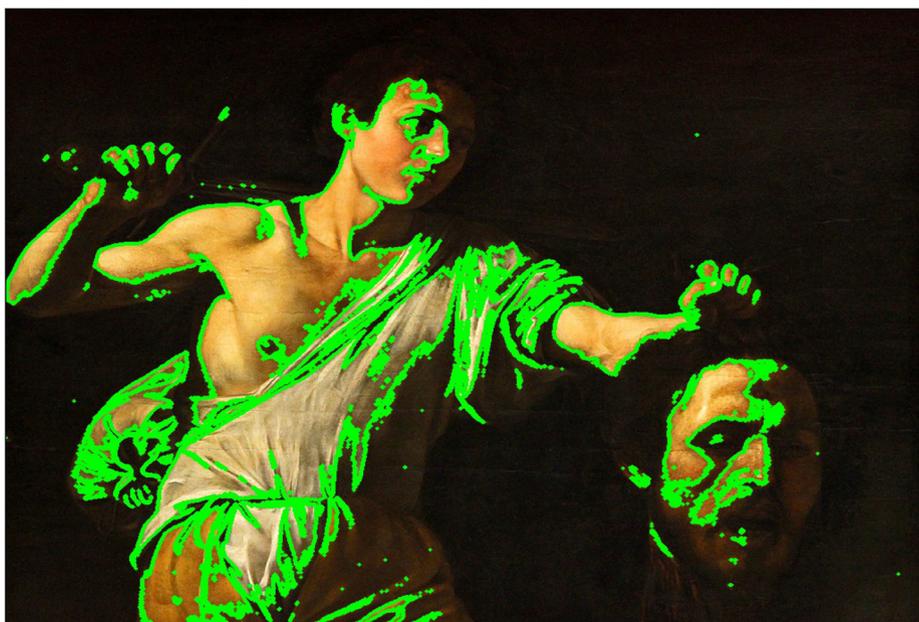


Figura 37. Detección de contornos de la pintura *David con la cabeza de Goliat* (1609-1610), de Michelangelo Merisi da Caravaggio



Figura 38. Detección de bordes de la pintura *La ciencia inútil o el alquimista* (1958), de Remedios Varo

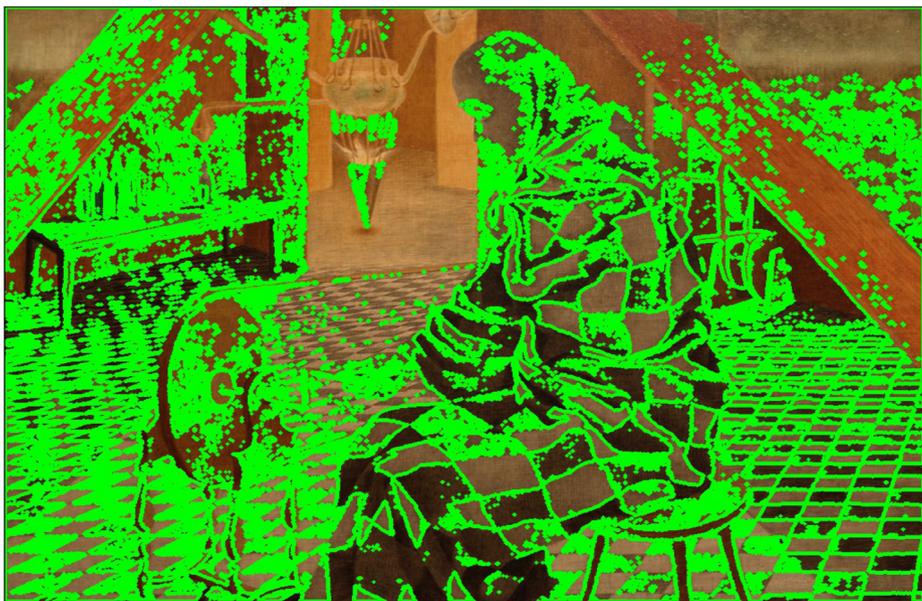


Figura 39. Detección de contornos de la pintura *La ciencia inútil o el alquimista* (1958), de Remedios Varo

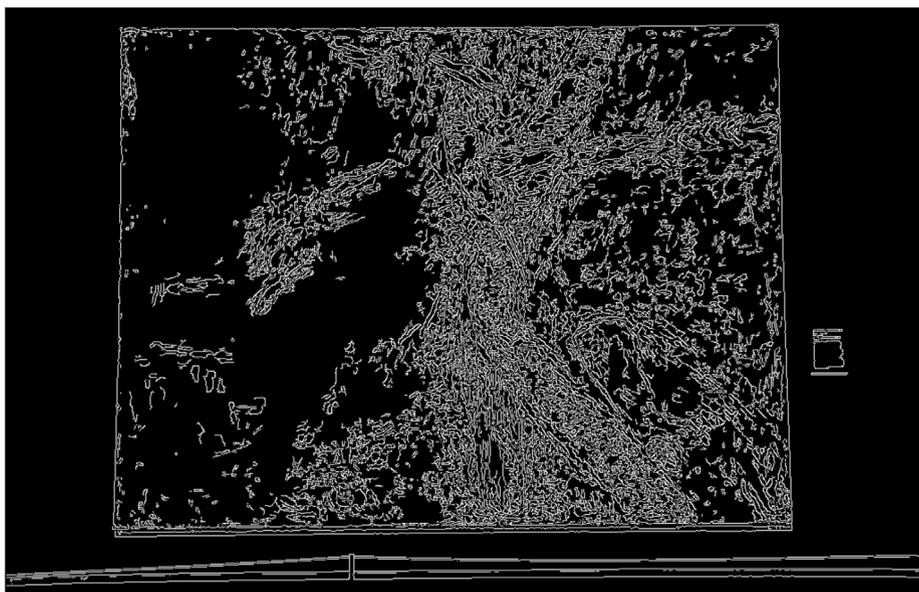


Figura 40. Detección de bordes de la pintura *Galisteo Creek* (1992), de Susan Rothenberg

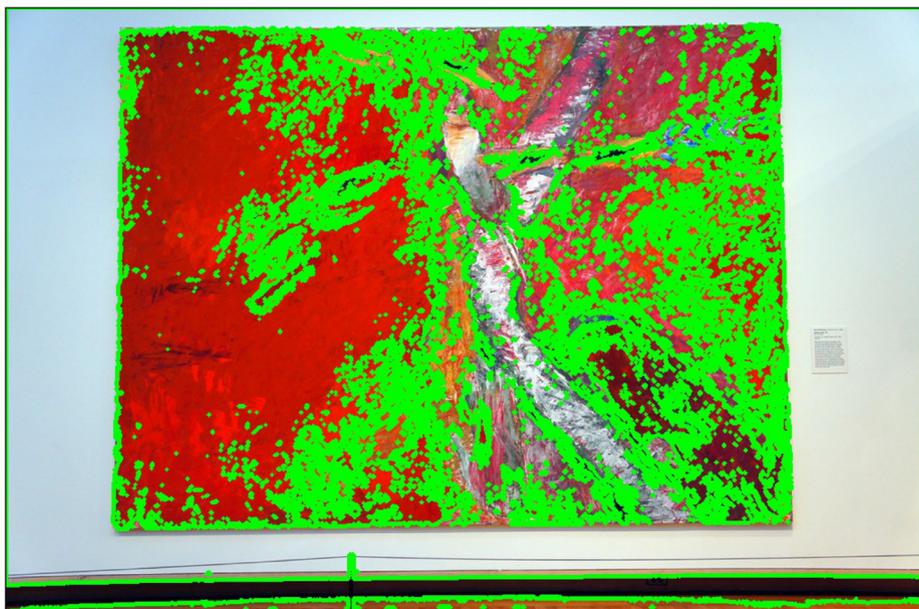


Figura 41. Detección de contornos de la pintura *Galisteo Creek* (1992), de Susan Rothenberg

El modelo de clasificación que fue utilizado durante este estudio permite identificar cinco tipos distintos de imágenes referentes a obras de las artes plásticas —dibujo, grabado, iconografía, pintura y escultura—; a través del filtrado, detección y extracción simultánea de sus atributos visuales más destacados, como lo son los aspectos cromáticos y lumínicos predominantes, su concentración o dispersión en secciones del espacio de la imagen, las formas y texturas presentes en la obra (Weinmann, 2013; Zunzunegui Díez, 2010), el modelo incorporado a la aplicación web<sup>20</sup> con la que culmina este estudio “devuelve” una estimación probabilística de la clase a la que pertenece la imagen. Para probar este modelo se emplearon por lo menos tres ejemplos diferentes de cada una de las clases establecidas; los resultados obtenidos fueron variados, dado que en ocasiones el sistema catalogó una imagen como correspondiente a otra clase, pese a que a la vista humana era evidente el tipo de obra.

Con respecto a la categoría de dibujo, el modelo logró identificar con una precisión aceptable los tres ejemplos brindados; la imagen que se muestra en la figura 42 el sistema la clasificó acertadamente como un dibujo con una probabilidad de 95%; de manera parecida, en lo que concierne a la categoría del grabado, el modelo tuvo también un desempeño suficiente, pues la estimación probabilística para cada caso osciló entre 50 y 70%, como en el caso de la figura 43, que se trata de un famoso grabado del pintor español Francisco de Goya.

Por su parte, ninguno de los ejemplos con los que se verificó la clase de obra denominada como iconografía obtuvo una estimación correcta; en el ejemplo que se muestra en la figura 44, el modelo clasificó a la imagen como un dibujo; en lo que respecta a la categoría de pintura, se tuvieron que emplear más ejemplos para comprobar el funcionamiento del sistema, dado que tuvo complicaciones para la clasificación correcta de las imágenes, catalogándolas como iconografías (ver figura 45) o como esculturas. En el último de los casos, el modelo no tuvo mayores complicaciones para identificar las imágenes con las que se probó como pertenecientes al tipo de esculturas.

<sup>20</sup> Se invita a los lectores y lectoras interesados e interesadas a probar por sí mismos la aplicación de visión computacional desarrollada a partir de este estudio accediendo al siguiente enlace: <https://share.streamlit.io/jeoa-1981/appiconology/main/src/AppIconology.py>

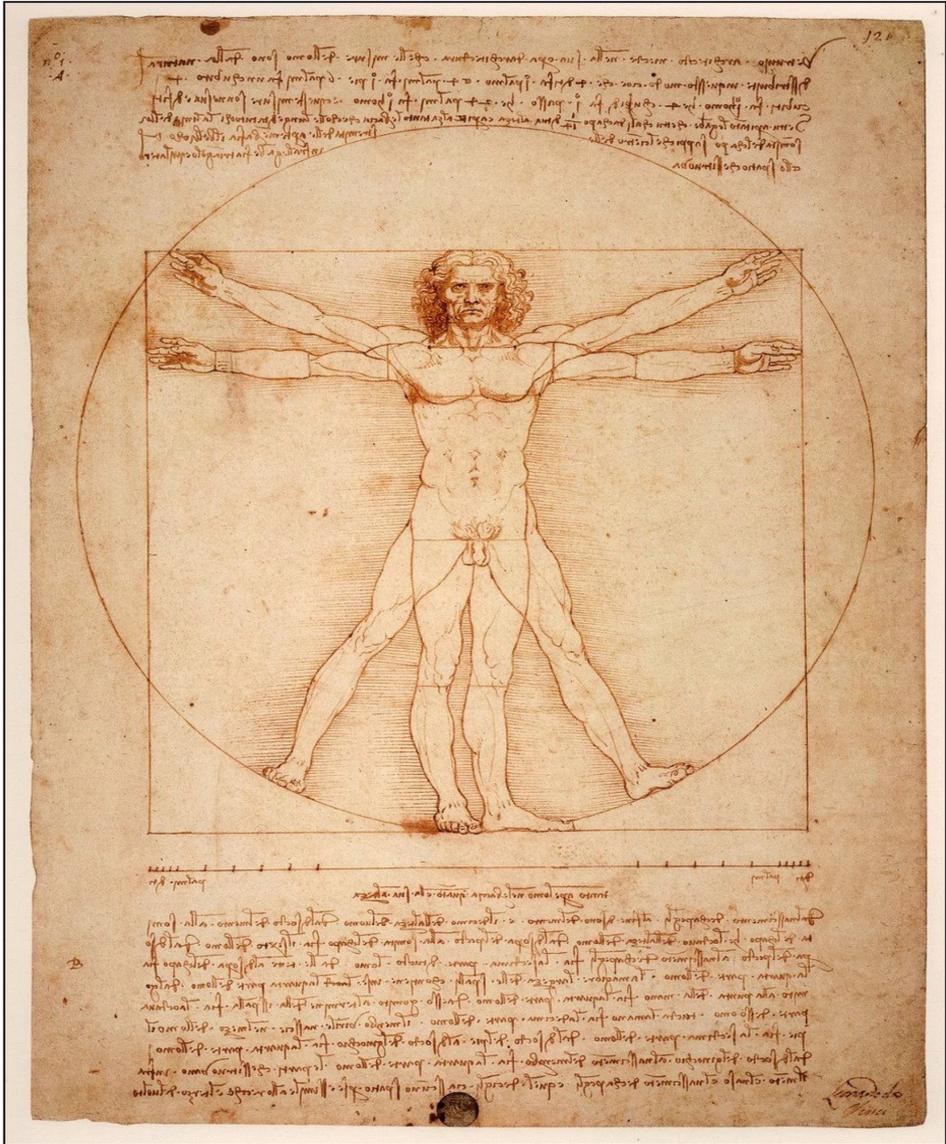


Figura 42. Leonardo da Vinci, *El hombre de Vitruvio*, 1492



Figura 43. Francisco de Goya,  
*El sueño de la razón produce monstruos*, 1799

Figura 44. Fachada de la parroquia  
San Fernando, Ciudad de México





Figura 45. María Izquierdo, *La tragedia*, 1949



Figura 46. Umberto Boccioni, *Formas únicas de continuidad en el espacio*, 1913

Después de esta sucinta aproximación al uso de técnicas de visión computacional para el análisis de imágenes de obras artísticas de la plástica, conviene, o bien reiterar algunas observaciones que han surgido a lo largo de este abordaje panorámico, o bien profundizar en torno a algunos aspectos que se han constatado anteriormente. En lo que respecta al uso de modelos y funciones para la identificación de objetos, se insiste en el gran potencial que tiene esta actividad para la localización automatizada y para el examen de motivos artísticos que posibiliten una ordenación, quizá más rápida y comprensiva, de obras de arte según periodos o estilos artísticos. Sobre la detección de bordes y contornos, su utilidad, se insiste, radica en que permite observar rasgos estilísticos propios de la obra, esto es, facilita el análisis de sus elementos esenciales desde una óptica que podría denominarse “hilemorfista”<sup>21</sup> —como una unidad de materia y forma— (Tatar-kiewicz, 2002).

Con respecto al modelo de clasificación incorporado a la aplicación web resultante de este estudio, es pertinente reiterar que se requiere un re-entrenamiento con documentos visuales —o imágenes— más diversificados para así incrementar las posibilidades de una correcta catalogación; cabe agregar, no obstante, que durante su prueba se pudo verificar que se requiere también de un procesamiento previo de las imágenes que compongan el corpus para eliminar o suprimir elementos que no sean pertinentes para el entrenamiento del modelo —por ejemplo, bordes de cuadros, suelos del espacio de exposición, fichas técnicas que puedan aparecer, entre otro tipo de elementos que podrían considerarse como externos o ajenos a la pieza contenida en la imagen—; aunado a lo anterior, el modelo fue entrenado y probado con imágenes de licencia abierta —que pueden reproducirse y reutilizarse libremente— que no fueron editadas para mejorar la identificación de sus atributos visuales, por lo que, a partir de una mejor preparación del corpus para el entrenamiento del modelo, se puede esperar un mejor desempeño para la clasificación de imágenes de obras de arte.

---

<sup>21</sup> Heidegger, en cierta medida y desde una óptica fenomenológica, ya había ensayado un análisis que pudiera considerarse basado en la materia y la forma del ser de la obra de arte (ver Heidegger, 1999).



## Conclusiones

Un concepto amplio de la cultura la definiría como el conjunto de modos y prácticas mediante las cuales producimos e intercambiamos sentido —ideas, valores, visiones de mundo—. Desde diversas perspectivas teóricas, lo mismo que desde varios planteamientos de políticas públicas, la cultura se presenta como una dimensión humana esencial y fundamental tanto para la vida en sociedad como para la realización personal. Si bien desde la década de 1980 gobiernos y empresas vuelven la vista hacia la cultura como una fuente de oportunidades económicas —para la generación de empleos y ocupación laboral, lo mismo que para la creación y adición de valor—, la tendencia no sólo se ha mantenido hasta hoy, sino que además se ha extendido y popularizado la idea del influjo positivo que la cultura ejerce en las dinámicas sociales.

En este contexto, las actividades culturales y creativas que pudieran definirse como aquellas orientadas a la producción de bienes de valor cultural generados a través de la creatividad, individual o grupal, constituyen la base sobre la cual se erige y se sustenta la tesis de la cultura como un recurso para el desarrollo, y de la creatividad como un insumo para la producción de bienes y servicios en general. Además, las prácticas clasificadas como culturales y creativas conforman un paradigma económico que se ha denominado economía creativa o “naranja” —o también cultural, dependiendo de la perspectiva teórica que se adopte—, que a su vez forma parte de la economía del conocimiento, centrada en la generación, uso y transformación

de datos e información en nuevos conocimientos aplicables a la mejora o creación de nuevos productos y servicios.

Por otra parte, y en concordancia con lo anterior, la digitalización, la interconectividad y las posibilidades de comunicación remota con las que contamos en la actualidad son fenómenos que, al posibilitar un acceso mayor a los bienes culturales, así como a herramientas para la colaboración creativa —aun sin coincidir físicamente en el mismo espacio— propician, según analistas y expertos, el crecimiento del sector cultural e incentivan la difusión de obras de arte, ejecuciones artísticas y otros proyectos culturales y creativos más allá de las fronteras geográficas. En otras palabras, las actividades culturales y creativas se han visto ampliamente beneficiadas por la convergencia tecnológica y comunicativa, no sólo en lo que respecta a las posibilidades de su difusión y promoción, sino también en lo que respecta a los procesos de creación, así como en las alternativas de comercialización y en la captación de una mayor audiencia.

Pese a todas las oportunidades y ventajas que brindan a la producción cultural y a la economía creativa la convergencia mediática y la digitalización, también plantean nuevas problemáticas y retos al sector (Flew, 2014; Hesmondhalgh, 2013), de entre los cuales se destacan, primero, el acceso limitado y desigual a internet, lo que condiciona tanto el acceso a bienes y servicios culturales como a la posibilidad de generar nuevos mensajes simbólicos a través de la red; segundo, si bien la digitalización y la convergencia mediática pueden conducir a una democratización y descentralización cultural, también pueden generar nuevas formas de exclusión y centralización en función de la población que es competente en el uso de las nuevas tecnologías de la información y la comunicación respecto de la que no posee el mismo grado de alfabetización digital; tercero, la piratería se ha convertido en una problemática cada vez más relevante en el contexto digital, afectando el reconocimiento autoral y la distribución legal de los contenidos culturales; finalmente, los diversos tratamientos, cada vez más sofisticados, de los enormes flujos de datos que circulan en la sociedad contemporánea (Ontiveros Baeza y López Sabater, 2018) impelen a sospechar que existen nuevas formas de vigilancia permanente, así como estrategias de dominación ideológica a través de la cultura (Hesmondhalgh, 2013; Lessig, 2005).

De acuerdo a lo anterior, cada uno de los estudios que conforman el proyecto presentado en este libro está orientado a brindar una serie de soluciones o, en su caso, acciones que contribuyan a la mejora y al desarrollo del sector cultural y creativo en la sociedad datificada, que se perfila o que es, más bien, una forma de estructuración socioeconómica fundada en fenómenos que hasta hace dos décadas pertenecían a la imaginaria de la ciencia ficción: (a) vehículos autónomos, manufactura aditiva o impresión 3D, robótica y androides operantes con inteligencia artificial cada vez más sofisticada, (b) nuevos modelos de negocios y de finanzas, como las criptomonedas o el paradigma *on-demand*, el internet de las cosas y (c) la tecnología protésica y el mejoramiento genético. Es evidente que la sociedad datificada supone el replanteamiento no sólo de la organización de los medios y modos de producción o de los modelos de negocios y las interacciones entre industria, academia y gobierno, sino también de las formas en que los individuos y las comunidades se expresan, se comunican y se interrelacionan.

Toda transformación o cambio de paradigma implica oportunidades y ventajas, lo mismo que riesgos y amenazas; la sociedad datificada no es una excepción a este hecho. De entre los múltiples beneficios que conlleva se destaca la transición de una economía de la propiedad a una economía de la colaboración, una ampliación de la oferta laboral y el potenciamiento de las capacidades humanas a través de las tecnologías emergentes. No obstante, y como se comentó anteriormente, la velocidad y magnitud con la que generamos, compartimos, intercambiamos y consumimos datos de diferentes clases es abrumadora, y puede derivar en una enorme dificultad para responder y adaptarnos, como sociedad y como individuos, a las necesidades que estos vertiginosos cambios nos plantean.

Ante este escenario, se requiere afrontar a la sociedad datificada desde una perspectiva humanista y humanizadora, lo que implica posicionar al ser humano como el eje de todos estos cambios, mas no en un sentido centralizador o antropocentrista, sino más bien con una orientación integradora. En este sentido, las artes y la cultura nos aprovisionan de las herramientas cognitivas y emocionales para afrontar los retos del porvenir desde una postura empática e incluyente. En otras palabras, “humanizar” a la sociedad datificada significa reconocer que el desarrollo tecnológico es sólo

una de las fuerzas motoras de esta reorganización socioeconómica, y que las interacciones humanas son la contraparte necesaria para lograr la creación de una sociedad equitativa, justa y sostenible. Aún más, las artes, la cultura y la creatividad constituyen el medio más adecuado para dotar de significación humana a toda transformación contemporánea.

Existe, además, la necesidad de repensar y reformular nociones, conceptos y teorías sobre el fenómeno artístico, la creatividad, la estética y la producción cultural en general. Dos de los atavismos vigentes aún en el sector cultura son, primero, el de la reproducción de la división de clases sociales en espectros culturales o, lo que es lo mismo, la diferenciación peyorativa entre alta cultura, cultura de masas y cultura popular y, segundo, el de la separación tajante entre las artes y la cultura respecto a las demás esferas de la praxis humana. Ambas premisas son igualmente erróneas, pues, en lo que respecta a la primera, desde una teoría ampliada de la cultura, la alta cultura, que comprende los artefactos artísticos de culto o más refinados (literatura clásica, obras pictóricas o piezas musicales de los grandes maestros, etc.), sólo se diferencia de la cultura de masas y de la cultura popular por las posibilidades de acceso y por la adquisición de códigos que permitan “disfrutar” una obra clásica, lo que implica una problemática educativa.<sup>22</sup>

En lo que respecta a la segunda premisa, las obras y las prácticas artísticas y culturales no existen aisladas de la política o de la economía, al contrario, constituyen objetos y procesos de intercambio simbólico, cuyo significado integral sólo puede ser captado en tanto se comprende toda la urdimbre social en la que están insertos. En el contexto de la sociedad daticada es necesario, si no es que urgente, repensar las relaciones de las artes con la sociedad, lo mismo que el sentido de la cultura como un recurso para el desarrollo integral del ser humano, además de reconsiderar las nuevas problemáticas que atañen a la creación y enseñanza de las artes en un mundo globalizado y que tiende a una digitalización abarcadora. En el caso particular de las profesiones creativas (artistas, diseñadores, literatos, escenógrafos, realizadores audiovisuales, actores, bailarines y músicos) es

---

<sup>22</sup> Y también de índole estética, en tanto existen diferencias cualitativas en la apreciación de la obra que no dependen únicamente del bagaje cultural o del capital cultural de la persona, sino también de los usos y funciones de los elementos propios con los que se confecciona un trabajo artístico, es decir, de sus caracteres formales.

necesario reflexionar en torno a su vigencia y pertinencia en un panorama dominado por las nuevas tecnologías de la información y la comunicación.

Ahora bien, en lo que concierne a la creación artística y creativa, el aprovechamiento de las tendencias y tecnologías actuales puede derivar tanto en innovaciones culturales, basadas en una articulación efectiva entre la capacidad creativa y las herramientas tecnológicas y digitales, como en una mayor difusión de las obras a través de las diferentes plataformas interconectadas disponibles (desde la computadora de escritorio hasta el dispositivo móvil). El uso de las tecnologías y modelos de producción de la sociedad datificada para la creación artística y creativa en general conlleva sin duda a la creación de nuevos objetos estéticos, lo mismo que a nuevas formas por medio de las cuales los públicos se relacionarán con obras con un alto grado de interactividad que conviertan a los espectadores en partícipes o co-creadores.

Con respecto a la articulación económica de las actividades culturales y creativas, éstas se pueden beneficiar ampliamente de las transformaciones que caracterizan a la sociedad datificada de tres maneras complementarias: (a) empleando las plataformas digitales en línea para la generación y circulación masiva de contenidos culturales; (b) expandiendo los mercados culturales y creativos aprovechando las herramientas de las diversas plataformas digitales (algoritmos de búsqueda, herramientas de gestión de contenidos, campañas virtuales) y (c) empleando *software* de código abierto (*Open Source*) y herramientas de la cultura *Maker* o *Do It Yourself* (DIY o “hazlo tú mismo”) para la creación de micro y pequeñas empresas culturales sostenibles y que establezcan sinergias entre sí para propiciar un crecimiento y desarrollo paulatinos. En suma, puede afirmarse que el sector cultural y creativo enfrenta el reto de dotar de significación humana a la sociedad datificada y, a un mismo tiempo, se beneficiaría profundamente de la apropiación de sus tecnologías, modelos y tendencias para la formación de profesionales creativos altamente competentes, y para la innovación cultural y creativa en general.

Así pues, y con base en los estudios que conforman este proyecto de analítica cultural, se concluye este libro presentando tres recomendaciones generales para un desarrollo del sector cultural y creativo asistido por diversas herramientas tecnológicas:

(a) *Para la investigación y la enseñanza de las artes y la cultura:* En este rubro se aconseja incorporar herramientas de las humanidades digitales y de la ciencia de datos para el análisis de las artes y demás expresiones culturales y creativas. Concretamente, en lo que concierne a la investigación en estos ámbitos, se recomienda el uso de *software* especializado para minería de datos, como KNIME, Weka, Orange, Voyant, entre otros; así mismo, se sugiere emplear estrategias y métodos correspondientes al NLP y del entendimiento del NLU, el raspado de red (Web Scraping), y el procesamiento de imágenes empleando visión computacional (Machine Vision), adaptados a las necesidades específicas de la investigación de las artes y la cultura.

En lo que respecta a la enseñanza de las artes, de igual manera se pueden emplear las estrategias y el *software* referido previamente para fomentar las capacidades analíticas del estudiantado, pero dando prioridad a la generación de objetos multimedia de aprendizaje por parte del mismo alumnado, empleando tecnologías tales como la realidad aumentada (*Augmented Reality*), la realidad virtual (*Virtual Reality*) y el internet de las cosas (*Internet of Things*). El doble propósito de promover un uso activo de estas tecnologías por parte del alumnado es, primero, fomentar el autoaprendizaje para la resolución creativa de problemas específicos y, segundo, impulsar la creatividad, la empatía y la sensibilidad a través del desarrollo de sus capacidades expresivas. Para ello, el mismo sector académico puede desarrollar MOOCs (*Masive Open Online Course*, curso en línea masivo y abierto) sobre el uso de las tecnologías previamente comentadas para la creación artística, la producción audiovisual y el diseño gráfico.

(b) *Del uso de las tecnologías actuales para la creación artística y la producción creativa:* En este eje se hace hincapié en la apropiación y el uso creativo de las tecnologías que conforman la sociedad datificada para la generación de proyectos artísticos y creativos que deriven en nuevos objetos estéticos multimedia e interactivos o aproximaciones innovadoras a expresiones artísticas clásicas y/o al patrimonio cultural consagrado. El término apropiación se refiere, en este caso en particular, a la exploración sistemática de los usos estanda-

rizados de las tecnologías actuales para, posteriormente, direccionarlas a la creación de obras o ejecuciones artísticas y otras prácticas creativas que empleen la realidad aumentada y la realidad virtual, la robótica y la manufactura aditiva como elementos sustantivos de su composición y que coadyuven, además, a su difusión.

Mediante el uso de las tecnologías, técnicas y métodos descritos en este libro, se puede impulsar la digitalización y restauración del patrimonio artístico de museos, galerías y archivos históricos; en este caso, la visión computacional posibilita la mejora de las imágenes, así como la creación de respaldos digitales de alta calidad de obra pictórica y gráfica, y reproducciones fidedignas de obras plásticas con impresión 3D. Se recomienda, además, replicar estos procedimientos para la protección y resguardo del patrimonio artístico y cultural de diversos organismos e instituciones, sean museos, galerías y/o empresas privadas, abriendo así un campo que permita una mayor interacción entre la academia y el sector gubernamental y privado.

- (c) *De la aplicación de las tecnologías actuales para el desarrollo y fomento económico de las actividades culturales y creativas:* Se sugiere la conformación tanto de clústeres interdisciplinarios, entendidos como grupos de trabajo interdisciplinarios que integren profesionales de diversas áreas para la realización de proyectos culturales y/o creativos que tengan impacto empresarial, así como de *startups* centradas en la producción y oferta de bienes culturales y creativos, y basadas en el uso sistemático y aplicación de las tecnologías más recientes de acceso abierto para la mejora e innovación continuas del ciclo productivo, propiciando además la escalabilidad y sostenibilidad de estas iniciativas emprendedoras. El esquema propuesto para este ecosistema es el de la integración horizontal de académicos, productores y estudiantes que, de manera conjunta y colaborativa, sean capaces de idear proyectos innovadores y/o modelos de negocios disruptivos de base cultural y creativa, apoyados en tecnologías, modelos y tendencias de la sociedad datificada, y que deriven en beneficios para la sociedad, así como en el fortalecimiento de los valores culturales.

Para finalizar, es importante mencionar que cada una de las recomendaciones o propuestas previamente explicadas implican no sólo la apropiación y conocimiento de una o varias tecnologías en particular ni mucho menos la pura ejecución de acciones específicas para la realización o cumplimiento de objetivos específicos, al contrario, remiten a una toma de postura: el posicionamiento tanto de los académicos, como de los profesionistas creativos ante todas las transformaciones que vienen aparejadas con la sociedad datificada debe ser siempre desde el compromiso con la generación de conocimiento, lo mismo que con la creación de símbolos por medio de los cuales se compartan visiones diversas. En este sentido, el análisis de las prácticas culturales y creativas contemporáneas ha de conducir, pues, a la democratización de los medios de producción de sentido o, en otras palabras, a que se incrementen y se mejoren las posibilidades de participación de cada persona en la producción cultural.

## Referencias

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X. (2016). TensorFlow: A System for Large-Scale Machine Learning. *OSDI'16: Proceedings of the 12<sup>th</sup> USENIX conference on Operating Systems Design and Implementation*, 265-283. Recuperado de <https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf>
- Ajitsaria, A. (s. f.). *Build a Recommendation Engine with Collaborative Filtering*. Recuperado el 7 de marzo de 2021 de <https://realpython.com/build-recommendation-engine-collaborative-filtering/>
- Álamo Felices, F. (2010). El microrrelato: Análisis, conformación y función de sus categorías narrativas. *Signa: Revista de la Asociación Española de Semiótica*, 19. <http://revistas.uned.es/index.php/signa/article/view/6233>
- Albon, C. (2018). *Machine Learning with Python Cookbook*. O'Reilly Media.
- Altamirano, C., y Sarlo, B. (1990). *Conceptos de sociología literaria*. Centro Editor de América Latina.
- Anandarajan, M., y Anandarajan, M. (Eds.). (2010). *e-Research Collaboration*. Berlín Heidelberg: Springer-Verlag. <https://link.springer.com/book/10.1007%2F978-3-642-12257-6>
- Arvidson, J., Askander, M., Bruhn, J., y Fuhrer, H. (Eds.). (2007). *Changing Borders: Contemporary Positions in Intermediality*. (Intermedia Studies Press; Vol. 1). Intermedia Studies Press.
- Baak, M., Koopman, R., Snoek, H., y Klous, S. (2019). A New Correlation Coefficient between Categorical, Ordinal and Interval Variables with Pearson Characteristics. *ArXiv:1811.11440 [Stat]*. <https://arxiv.org/abs/1811.11440>
- Ballard, W. (2018). *Hands-On Deep Learning for Images with TensorFlow: Build Intelligent Computer Vision Applications Using TensorFlow and Keras*. Packt.
- Barthes, R., Lefebvre, H., y Goldmann, L. (1969). *Literatura y sociedad. Problemas de metodología en sociología de la literatura*. Ediciones Martínez Roca.
- Bauman, Z. (2002). *La cultura como praxis*. Paidós.
- Becker, H. S. (2008). *Los mundos del arte. Sociología del trabajo artístico*. Universidad Nacional de Quilmes.
- Bernal, C. A. (2010). *Metodología de la investigación: Administración, Economía, humanidades y ciencias sociales*. Pearson Educación.
- Berry, D. M., y Fagerjord, A. (2017). *Digital humanities: Knowledge and critique in a digital age*. Polity.
- Beysolow II, T. (2018). *Applied Natural Language Processing with Python: Implementing Machine Learning and Deep Learning Algorithms for Natural Language Processing*. Apress. <https://link.springer.com/book/10.1007%2F978-1-4842-3733-5>
- Bird, S., Klein, E., y Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly. <http://www.datascienceassn.org/sites/default/files/Natural%20Language%20Processing%20with%20Python.pdf>

- Bocanegra Barbecho, L., y García López, A. (Eds.). (2017). *Con la red/en la red: Creación, investigación y comunicación cultural y artística en la era Internet*. Downhill Publishing.
- Bourdieu, P. (2002a). *Campo de poder, campo intelectual. Itinerario de un concepto*. Montessor.
- Briggs, A., y Burke, P. (2005). *De Gutenberg a internet: Una historia social de los medios de comunicación*. Taurus.
- Burdick, A. (Ed.). (2012). *Digital humanities*. MIT Press.
- Calderón, C. A. (2013). *e-Investigación en Ciencias Sociales: Adopción y uso de TIC por investigadores sociales de América Latina*. CLACSO.
- Campra, R. (2008). *Territorios de la ficción: Lo fantástico*. Renacimiento.
- Carroll, L. (2005). Capítulo X: La contradanza de los bogavantes. En *Aventuras de Alicia en el País de las Maravillas* (p. 116). Akal Ediciones.
- Castells, M. (2000). *La era de la información: Economía, sociedad y cultura*. Alianza.
- Chen, H., Wang, Y., y Cheung, K.-H. (Eds.). (2010). *Semantic e-Science* (Vol. 11). Springer. <https://link.springer.com/book/10.1007/978-1-4419-5908-9>
- Colbert, F., y Cuadrado, M. (2015). *Marketing de las artes y la cultura*. Ariel.
- Creeber, G. (Ed.). (2009). *Digital cultures: Understanding new media*. Open University Press.
- Creswell, J. W. (2014). *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.
- Cros, E. (1986). *Literatura, ideología y sociedad*. Gredos.
- Cuche, D. (2002). *La noción de cultura en las ciencias sociales*. Nueva Visión.
- Dawson-Howe, K. (2014). *A practical introduction to computer vision with OpenCV*. John Wiley & Sons.
- Debray, R. (2010). *Vida y muerte de la imagen: Historia de la mirada en Occidente*. Paidós.
- Drucker, J., Kim, D., Salehian, I., y Bushong, A. (2014). *Introduction to Digital Humanities*. UCLA.
- Dwivedi, R. (2020). *Introduction to Natural Language Processing: Text Cleaning & Preprocessing | Analytics Steps*. <https://www.analyticssteps.com/blogs/introduction-natural-language-processing-text-cleaning-preprocessing>
- Eco, U. (2000). *Tratado de semiótica general*. Lumen.
- Elleström, L. (Ed.). (2010). *Media borders, multimodality and intermediality*. Palgrave Macmillan.
- Embarak, Dr. O. (2018). *Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems*. Apress. <https://link.springer.com/book/10.1007%2F978-1-4842-4109-7>
- Escarpit, R. (1980). *Sociología de la literatura*. oikos-tau.
- Estalella, A., y Ardévol, E. (2011, enero-abril). e-research: desafíos y oportunidades para las ciencias sociales. *Convergencia. Revista de Ciencias Sociales*, 18(55), 1405-1435. <https://www.redalyc.org/pdf/105/10515210004.pdf>
- Flew, T. (2014). *New Media*. Oxford University Press.
- Foxwell, H. J. (2020). *Creating Good Data: A Guide to Dataset Structure and Data Representation*. Apress. <https://link.springer.com/book/10.1007%2F978-1-4842-6103-3>

- Gander, J. (2017). *Strategic Analysis. A Creative and Cultural Industries Perspective*. Routledge.
- García Canclini, N. (2014). *La producción simbólica. Teoría y método en sociología del arte*. Siglo XXI Editores.
- García Herrero, J., Berlanga de Jesús, A., Molina López, J. M., Patricio Guisado, M. Á., Bustamante, Á. L., y Padilla, W. R. (2018). *Ciencia de datos. Técnicas analíticas y aprendizaje estadístico*. Alfaomega.
- García Zaballos, A., y Iglesias Rodríguez, E. (2017). *Economía digital en América Latina y el Caribe: Situación actual y recomendaciones*. Inter-American Development Bank. <https://publications.iadb.org/en/economia-digital-en-america-latina-y-el-caribe-situacion-actual-y-recomendaciones>
- Gardner, H. J., y Manduchi, G. (2007). *Design patterns for e-science*. Springer.
- Garrido Domínguez, A. (Ed.). (1997). *Teorías de la ficción literaria*. Arco/Libros.
- Gollapudi, S. (2019). *Lean Computer Vision Using OpenCV: With Deep Learning CNNs and RNNs*. Apress.
- Gonçalves, L. (2020, noviembre 5). *Automatic Text Summarization with Machine Learning — An Overview*. Medium. <https://medium.com/luisfredgs/automatic-text-summarization-with-machine-learning-an-overview-68ded5717a25>
- Goriunova, O. (2012). *Art Platforms and Cultural Production on the Internet*. Routledge.
- Gronlund, M. (2017). *Contemporary Art and Digital Culture*. Routledge.
- Gubern, R. (1996). *Del bisonte a la realidad virtual: La escena y el laberinto*. Anagrama.
- Hartley, J., Wen, W., y Siling Li, H. (2015). *Creative Economy and Culture*. Sage Publications.
- Heidegger, M. (1999). *Arte y poesía*. Fondo de Cultura Económica de España.
- Heinich, N. (2002). *Sociología del arte*. Nueva Visión.
- Henley, A. J., y Wolf, D. (2018). *Learn Data Analysis with Python*. Apress. <https://link.springer.com/book/10.1007%2F978-1-4842-3486-0>
- Hernández Sampieri, R., Fernández Collado, C., y Baptista Lucio, M. del P. (2014). *Metodología de la investigación*. McGraw-Hill.
- Herzogenrath, B. (Ed.). (2012). *Travels in intermedia[lity]: Reblurring the boundaries*. Dartmouth College Press.
- Hesmondhalgh, D. (2013). *The Cultural Industries*. Sage Publications.
- Hine, C. (Ed.). (2006). *New infrastructures for knowledge production: Understanding E-science*. Information Science Pub.
- Hine, C. (2008). *Systematics as cyberscience: Computers, change, and continuity in science*. MIT Press.
- Hine, C. (2015). *Ethnography for the Internet: Embedded, Embodied and Everyday*. Routledge.
- Honnibal, M., Montani, I., Van Landeghem, S., y Boyd, A. (2020). *spaCy: Industrial-strength Natural Language Processing in Python*. Zenodo. <https://doi.org/10.5281/zenodo.1212303>
- Howse, J., Joshi, P., y Beyeler, M. (2016). *OpenCV: Computer vision projects with Python: get savvy with OpenCV and actualize cool computer vision applications: a course in three modules*. Packt.

- Hug, N. (2020, agosto 5). Surprise: A Python library for recommender systems. *The Journal of Open Source Software*, 5(52), 2174. <https://joss.theoj.org/papers/10.21105/joss.02174>
- Jenkins, H. (2006). *Fans, bloggers, and gamers: Exploring participatory culture*. New York University Press.
- Jenkins, H. (2008). *Convergence Culture. La cultura de la convergencia de los medios de comunicación*. Paidós.
- Jenkins, H., Ford, S., y Green, J. (2015). *Cultura transmedia. La creación de contenido y valor en una cultura en red*. Gedisa.
- Joly, M. (2012). *Introducción al análisis de la imagen*. La Marca Editora.
- Jones, S. E. (2014). *The emergence of the digital humanities*. Routledge.
- Jullier, L. (2004). *La imagen digital: De la tecnología a la estética*. La marca.
- Karpathy, A. (2015, octubre 25). *What a Deep Neural Network thinks about your #selfie*. <http://karpathy.github.io/2015/10/25/selfie/>
- Kozinets, R. V. (2010). *Netnography: Ethnographic research in the age of the internet*. Sage Publications Ltd.
- Kulkarni, A., y Shivananda, A. (2019). *Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python*. Apress. <https://link.springer.com/book/10.1007%2F978-1-4842-4267-4>
- Kumar, S. (2021). *Develop a Language Translator System in Python*. Towards Data Science. <https://towardsdatascience.com/develop-a-text-language-detector-and-translator-system-in-python-6537d6244b10>
- Lagmanovich, D. (2006). *El microrrelato: Teoría e historia*. Menoscuarto.
- Landow, G. P. (2008). *Hipertexto 2.0: La teoría crítica y los nuevos medios en una época de globalización*. Paidós.
- Lane, H., Cole, H., y Max Hapke, H. (2019). *Natural Language Processing in Action. Understanding, analyzing, and generating text with Python*. Manning Publications.
- László Hajba, G. (2018). *Website Scraping with Python*. Apress.
- Lessig, L. (2005). *Cultura libre: Cómo los grandes medios usan la tecnología y las leyes para encerrar la cultura y controlar la creatividad*. LOM : Derechos Digitales.
- Lévy, P. (2013). The Creative Conversation of Collective Intelligence. En P. Aronoff y H. Scott (Trads.), *The participatory cultures handbook* (pp. 99-108). Routledge.
- Li, S. (2020, agosto 2). *Multi Class Text Classification With Deep Learning Using BERT*. Medium. <https://towardsdatascience.com/multi-class-text-classification-with-deep-learning-using-bert-b59ca2f5c613>
- Li, F., Zhou, Y., y Cai, T. (2019, noviembre 7). Trails of Data: Three Cases for Collecting Web Information for Social Science Research. *Social Science Computer Review*. doi.org/10.1177/0894439319886019
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., y Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *Lecture Notes in Computer Science*, 9905, 21-37. [https://link.springer.com/chapter/10.1007/978-3-319-46448-0\\_2](https://link.springer.com/chapter/10.1007/978-3-319-46448-0_2)
- Lope.ai. (2019, marzo 27). *Recommendation Systems from Scratch in Python | PYTHON*

- LABS. Medium. <https://medium.com/@lope.ai/recommendation-systems-from-scratch-in-python-pytholabs-6946491e76c2>
- Luzardo, A., Funes, G., y 23 Design. (2019). *Emprendimientos tecnocreativos: Creatividad y tecnología, ¿aliados o enemigos?* Inter-American Development Bank. [https://publications.iadb.org/publications/spanish/document/Emprendimientos\\_Tecnocreativos\\_Creatividad\\_y\\_tecnolog%C3%ADa\\_aliados\\_o\\_enemigos\\_es.pdf](https://publications.iadb.org/publications/spanish/document/Emprendimientos_Tecnocreativos_Creatividad_y_tecnolog%C3%ADa_aliados_o_enemigos_es.pdf)
- Lyons, J. (1984). *Introducción al lenguaje y a la lingüística*. Teide.
- Maanen, H. van. (2009). *How to study art worlds: On the societal functioning of aesthetic values*. Amsterdam University Press.
- Malik, U. (s. f.). *Creating a Simple Recommender System in Python using Pandas*. Stack Abuse. Recuperado el 7 de marzo de 2021 de <https://stackabuse.com/creating-a-simple-recommender-system-in-python-using-pandas/>
- Manovich, L. (2020). *Cultural analytics*. The MIT Press.
- Maria Farinella, G., Battiato, S., y Cipolla, R. (Eds.). (2013). *Advanced topics in computer vision*. Springer.
- Martin, T. (2017, marzo 23). *FiveThirtyEight*. <https://fivethirtyeight.com/features/dissecting-trumps-most-rabid-online-following/>
- Martínez Bonati, F. (2001). *La ficción narrativa: Su lógica y ontología*. LOM Ed.
- Mathur, P. (2019). *Machine Learning Applications Using Python. Cases Studies from Healthcare, Retail, and Finance*. Apress.
- Mattelart, A. (2002). *Historia de la sociedad de la información*. Paidós.
- McKinney, W. (2018). *Python for Data Analysis*. O'Reilly.
- McLuhan, M. (2007). *Comprender los medios de comunicación: Las extensiones del ser humano*. Paidós.
- McRobbie, A. (2016). *Be Creative. Making a Living in the New Culture Industries*. Polity Press.
- Mitchell, R. (2018). *Web Scraping with Python*. O'Reilly.
- Mittell, J. (2013). Wikis and Participatory Fandom. En *The participatory cultures handbook* (pp. 36-42). Routledge.
- Moro, S., Batista, F., Rita, P., Oliveira, C., & Ribeiro, R. (2019, septiembre 1). Are the States United? An Analysis of U.S. Hotels' Offers Through TripAdvisor's Eyes. *Journal of Hospitality & Tourism Research*, 43(7), 1112-1129. <https://doi.org/10.1177/1096348019854793>
- Moulin, R. (2012). *El mercado del arte: Mundialización y nuevas tecnologías*. La Marca.
- Nelli, F. (2018). *Python Data Analytics*. Apress.
- Olivé, L. (2007). *La ciencia y la tecnología en la sociedad del conocimiento: Ética, política y epistemología*. Fondo de Cultura Económica.
- Ontiveros Baeza, E., y López Sabater, V. (2018). *Economía de los datos*. Fundación Telefónica: Ariel.
- Panofsky, E. (2011). *El significado en las artes visuales*. Alianza.
- Patil, P. (2018, mayo 23). *What is Exploratory Data Analysis?* Medium. <https://towardsdatascience.com/exploratory-data-analysis-8fc1cb20fd15>

- Peist, N. (2012). *El éxito en el arte moderno: Trayectorias artísticas y proceso de reconocimiento*. Abada.
- Peng, R. D., y Matsui, E. (2016). *The Art of Data Science. A Guide for Anyone who Works with Data*. lulu.com.
- Perelman, C., y Olbrechts-Tyteca, L. (1989). *Tratado de la argumentación: La nueva retórica*. Gredos.
- Pink, S., Horst, H. A., Postill, J., Hjorth, L., Lewis, T., y Tacchi, J. (Eds.). (2016). *Digital ethnography: Principles and practice*. Sage.
- PNUD. (2014). *Informe sobre la economía creativa*. UNESCO.
- Poli, F. (1976). *Producción artística y mercado*. Gustavo Gili.
- Potter, J. (1998). *La representación de la realidad: Discurso, retórica y construcción social*. Paidós.
- Radim, R., y Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. En *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45-50). ELRA.
- Rafaeli, A., Ashtar, S., & Altman, D. (2019, agosto 14). Digital Traces: New Data, Resources, and Tools for Psychological-Science Research. *Current Directions in Psychological Science*, 28(6), 560-566. <https://doi.org/10.1177/0963721419861410>
- Raj, S. (2019). *Building Chatbots with Python. Using Natural Language Processing and Machine Learning*. Apress.
- Romero Frías, E., y Sánchez González, M. (Eds.). (2014). *Ciencias Sociales y Humanidades Digitales. Técnicas, herramientas y experiencias de e-Researche investigación en colaboración*. Sociedad Latina de Comunicación Social.
- Rowan, Jarod. (2010). *Emprendizajes en cultura. Discursos, instituciones y contradicciones de la empresarialidad cultural*. Traficantes de Sueños.
- Russell, M. A., y Klassen, M. (2019). *Mining the social web: Data mining Facebook, Twitter, LinkedIn, Instagram, Github, and more*. O'Reilly.
- Sarkar, D. (2019). *Text Analytics with Python*. Apress.
- Schirato, T., y Webb, J. (2004). *Reading the visual*. Allen & Unwin.
- Schmidt, S. J. (1997). La auténtica ficción es que la realidad existe. Modelo constructivista de la realidad, la ficción y la literatura. En A. Garrido Domínguez (Ed.), *Teorías de la ficción literaria*. Arco/Libros.
- Schwab, K. (2018). *Shaping the Future of the Fourth Industrial Revolution*. Currency.
- Science, O.-O. D. (2020, abril 2). *Creating a Data Analysis Pipeline in Python*. ODSC - Open Data Science. <https://medium.com/@ODSC/creating-a-data-analysis-pipeline-in-python-c7d434ce16e2>
- Scolari, C. A. (2013). *Narrativas transmedia: Cuando todos los medios cuentan*. Deusto. <https://www.overdrive.com/media/2011736/narrativas-transmedia>
- Sieg, A. (2018, julio 4). *Text Similarities: Estimate the degree of similarity between two texts*. Medium. <https://medium.com/@adriensieg/text-similarities-da019229c894>
- Siles, G. (2007). *El microrrelato hispanoamericano: La formación de un género en el siglo xx*. Corregidor.
- Smithies, J. (2017). *The Digital Humanities and the Digital Modern*. Palgrave Macmillan. <https://link.springer.com/book/10.1057%2F978-1-137-49944-8#toc>

- Spracklen, K. (2015). *Digital Leisure, the Internet and Popular Culture*. Palgrave MacMillan.
- Stepanek, H. (2020). *Thinking in Pandas: How to Use the Python Data Analysis Library the Right Way*. Apress. <https://link.springer.com/book/10.1007%2F978-1-4842-5839-2>
- Stoneman, P. (2010). *Soft Innovation. Economics, Product Aesthetics, and the Creative Industries*. Oxford University Press.
- Szurmuk, M., Irwin, R. M., Rabinovich, S., y Instituto de Investigaciones Dr. José María Luis Mora (Eds.). (2009). *Diccionario de estudios culturales latinoamericanos* (1. ed). Instituto Mora : Siglo Ventiuno Editores.
- Tandel, A. (2017, noviembre 1). *A Data Science Workflow*. Medium. <https://towardsdatascience.com/a-data-science-workflow-26c3f05a010e>
- Tatarkiewicz, W. (2002). *Historia de seis ideas: Arte, belleza, forma, creatividad, mimesis, experiencia estética*. Editorial Tecnos.
- Terzo, O., y Mossucca, L. (Eds.) (2017). *Cloud Computing with e-Science Applications*. CRC Press. <https://www.taylorfrancis.com/books/edit/10.1201/b18021/cloud-computing-science-applications-olivier-terzo-lorenzo-mossucca>
- The pandas development team. (2020). *pandas-dev/pandas: Pandas 1.2.4*. <https://zenodo.org/record/4681666#.YJJ-B2RKiRQ>
- Throsby, D. (2008). *Economía y cultura*. Conaculta.
- Towse, R. (2013). *A Textbook of Cultural Economics*. Cambridge University Press.
- vanden Broucke, S., y Baesens, B. (2018). *Practical Web Scraping for Data Science*. Apress. <https://link.springer.com/book/10.1007%2F978-1-4842-3582-9#about>
- Vasiliev, Y. (2020). *Natural language processing with Python and spaCy: A practical introduction*. No Starch Press.
- Verma, U. (2020, febrero 26). *Text Preprocessing for NLP (Natural Language Processing), Beginners to Master*. Analytics Vidhya. <https://medium.com/analytics-vidhya/text-preprocessing-for-nlp-natural-language-processing-beginners-to-master-fd82dfecf95>
- Weinmann, M. (2013). Visual Features — From Early Concepts to Modern Computer Vision. En G. Maria Farinella, S. Battiato, y R. Cipolla (Eds.), *Advanced Topics in Computer Vision* (pp. 1-34). Springer.
- Yang, X., Wang, L., y Jie, W. (Eds.). (2011). *Guide to e-Science*. Springer London. <https://link.springer.com/book/10.1007%2F978-0-85729-439-5>
- Yproductions. (2009). *Innovación en cultura. Una aproximación crítica a la genealogía y usos del concepto*. Traficantes de Sueños.
- Yúdice, G. (2002). *El recurso de la cultura*. Gedisa.
- Zallo, R. (1992). *El mercado de la cultura*. Gakoa.
- Zamora Águila, F. (2007). *Filosofía de la imagen: Lenguaje, imagen y representación*. UNAM, Escuela Nacional de Artes Plásticas.
- Zunzunegui Díez, S. (2010). *Pensar la imagen*. Cátedra : Universidad del País Vasco.
- Zychlinski, S. (2018, febrero 23). *The Search for Categorical Correlation*. towards data science. <https://towardsdatascience.com/the-search-for-categorical-correlation-a1cf7f1888c9>

*Cultura y Big Data. Métodos y técnicas  
para el análisis cultural en una sociedad dati-  
ficada*, de Jesús Eduardo Oliva Abarca, publicado  
por Ediciones Comunicación Científica S. A. de C. V.,  
se terminó de imprimir en mayo en los talleres de Ultradi-  
gital Press S. A. de C. V. Centeno 195, Col. Valle del Sur, 09819,  
Ciudad de México. El tiraje fue de 500 ejemplares impresos en cultural  
ahuesado de 75 gr



