

RICHARD ROGERS



MÉTODOS DIGITALES

IX-NIC IRUEGAS
TRADUCTORA



ITESO, Universidad
Jesuita de Guadalajara

MÉTODOS DIGITALES

RICHARD ROGERS



MÉTODOS DIGITALES

IX-NIC IRUEGAS
TRADUCTORA



ITESO, Universidad
Jesuita de Guadalajara

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE

Biblioteca Dr. Jorge Villalobos Padilla, S.J.

Rogers, Richard

Métodos digitales / R. Rogers ; tr. por Ix-Nic Iruegas. -- Guadalajara, México : ITESO, 2023.
375 p.

Traducción de: Digital Methods

ISBN 978-607-8910-04-5

1. Google – Aspectos Sociales y Culturales. 2. Buscadores de Información – Aspectos Sociales y Culturales. 3. Páginas en Línea – Aspectos Sociales y Culturales. 4. Redes Sociales (Internet) – Aspectos Sociales y Culturales. 5. Medios Sociales (Internet) – Aspectos Sociales y Culturales. 6. Internet – Aspectos Sociales y Culturales – Tema Principal. 7. Tecnologías de Comunicación e Información – Aspectos Sociales y Culturales – Tema Principal. 8. Comunicación Masiva. 9. Tecnología y Sociedad. 10. Sociología de la Comunicación. I. t.

[LC]

302. 23701 [Dewey]

Título original en inglés: *Digital Methods*

Traductora: Ix-Nic Iruegas

Diseño original: Danilo Design

Diseño de portada: Beatriz Díaz Corona J.

Diagramación: Beatriz Díaz Corona J.

Corrección: Brígida Botello

Foto de contraportada: Sabine Niederer

ITESO, Universidad Jesuita de Guadalajara

Dr. Alexander Paul Zatyryka Pacheco, S.J.

Rector

Dra. Catalina Morfín Otero

Directora General Académica

Dr. Humberto Orozco Barba

Director de Relaciones Externas

Mtro. Enrique Páez Agraz

Director del Departamento de Estudios Socioculturales

Dra. Rossana Reguillo

Coordinadora de Signa_Lab ITESO

Lic. Manuel Verduzco Espinoza

Director de la Oficina de Publicaciones

Mtra. María Guadalupe López Garfias

Coordinadora editorial

Original English language edition published by The MIT Press as Digital Methods by Richard Rogers.
©2013 Massachusetts Institute of Technology.

La edición original en inglés fue publicada por The MIT Press como *Digital Methods* por Richard Rogers.
©2013 Massachusetts Institute of Technology.

1a. edición, Guadalajara, 2023.

DR © Instituto Tecnológico y de Estudios Superiores de Occidente (ITESO)
Periférico Sur Manuel Gómez Morán 8585, Col. ITESO
Tlaquepaque, Jalisco, México, CP 45604
publicaciones.iteso.mx

Queda prohibida la reproducción parcial o total, directa o indirecta del contenido de la presente obra, sin contar previamente con la autorización expresa y por escrito de los editores, en términos de la Ley Federal del Derecho de Autor y, en su caso, de los tratados internacionales aplicables.

ISBN 978-607-8910-04-5

Impreso y hecho en México

Printed and made in Mexico

Índice

INTRODUCCIÓN: SITUACIÓN DE LOS MÉTODOS DIGITALES	7
1. EL FIN DE LO VIRTUAL: MÉTODOS DIGITALES	35
2. EL VÍNCULO Y LAS POLÍTICAS DEL ESPACIO WEB	69
3. EL SITIO WEB COMO OBJETO ARCHIVADO	101
4. LA GOOGLIZACIÓN Y EL MOTOR INOCENTE	133
5. LA BÚSQUEDA COMO INVESTIGACIÓN: LA DISTANCIA ENTRE FUENTES Y EL ANÁLISIS INTERESFÉRICO	151
6. ESTUDIOS SOBRE WEBS NACIONALES	195
7. LAS REDES SOCIALES Y LAS CARACTERÍSTICAS POSDEMOGRÁFICAS	237
8. WIKIPEDIA COMO REFERENCIA CULTURAL	253
9. DESPUÉS DEL CIBERESPACIO: MACRODATOS Y MICRODATOS	311

BIBLIOGRAFÍA

327

ÍNDICE ANALÍTICO

363

Introducción: situación de los métodos digitales

Este no es un libro de metodología, al menos no si se refiere con ello a un conjunto de técnicas y heurísticas que debe ser llevado de un lado a otro como si se tratara de una pesada caja de herramientas útiles en una gran variedad de ámbitos de investigación. Tampoco es el ejemplar más novedoso del manual de instrucciones ni una lista de respuestas a preguntas frecuentes, o algo que pudiera describir cómo operar software multipropósito a través del cual cierta cantidad de análisis estadísticos y de sistemas puede realizarse una vez que se ha recogido o entregado un conjunto de datos de forma separada. En cambio, este libro sí presenta una mirada metodológica para la investigación en la web. Como tal, es una propuesta para reorientar el campo de las investigaciones dirigidas a internet mediante el estudio y la reutilización de lo que llamo los métodos del medio, o quizá con mayor precisión, los métodos integrados en los dispositivos en línea; por ejemplo, si bien el rastreo, la extracción o *scraping*, la colaboración abierta distribuida y la *folcsonomía* son diferentes géneros y especies, en todos los casos se trata de técnicas de recolección y clasificación de datos en la web. PageRank y otros algoritmos similares son modos de ordenar y clasificar. Las nubes de palabras y otras formas comunes de visualización muestran la relevancia y la resonancia.

¿Cómo podemos aprender de y reutilizar estos y otros métodos en línea? El propósito no es tanto contribuir a sus ajustes y a la construcción de un mejor motor de búsqueda, ya que esa labor se optimiza si está en manos de las ciencias de la computación y las disciplinas semejantes. Más bien, la idea es, primero, pensar en conjunto con estas disciplinas y aprender cómo manejan los hipervínculos, las visitas, los *likes*, las etiquetas, los fechadores y otros objetos digitales nati-

vos. Al pensar siempre de la mano de los dispositivos y los objetos que manejan, los métodos digitales, como prácticas de investigación, aspiran a seguir los cambiantes métodos del medio.

Segundo, los métodos digitales no solo piensan con dispositivos en línea. También estudian la situación de la disponibilidad y el aprovechamiento de los objetos digitales para recombinarlos de la manera más productiva. Cuando se estudia un dispositivo web, se construye una nueva herramienta o se crea una interfaz por encima de una ya existente. La tarea consiste en registrar los elementos a nuestra disposición, por ejemplo, tuits, retuits, *hashtags*, nombres de usuario, ubicación de usuarios, URL recortados, respuestas arrobadas, etcétera (en el caso de Twitter, la plataforma de microblogueo). ¿Cómo pueden combinarse y recombinarse los objetos digitales de forma que sean útiles no tanto para la búsqueda en Twitter, sino para responder preguntas de investigación social y cultural? ¿Algún hashtag particular y su conjunto de tuits más retuiteados ofrecen una narración atractiva y organizada de algún evento y de quién se trata?

El tercer principio consiste en construir sobre los dispositivos dominantes existentes y con ellos llevar a cabo un diagnóstico cultural y social. Los métodos digitales reutilizan o construyen sobre los dispositivos dominantes del medio y, al hacerlo, crean trabajos que se derivan de los resultados, figurativa y literalmente; es decir, los resultados iniciales pueden ser los mismos o similares a los surgidos de los dispositivos en línea, pero se observan o se representan bajo una luz distinta, y convierten lo que alguna vez fue familiar —la página de resultados de un motor de búsqueda, una lista de tuits en orden cronológico inverso, una colección de comentarios o un conjunto de intereses obtenido de un perfil en redes sociales— en indicadores y conclusiones.

Las fuentes obtienen una clasificación alta en las páginas de resultados de los motores de búsqueda no solo porque son útiles a la hora de ofrecer información al usuario sobre la consulta que hicieron. La clasificación también se basa en el extenso análisis de los vínculos, la proporción de clics, la frescura y la observación lingüística, textual y geográfica que pueden ser revisados a través de códigos cualitativos que se concentran en una pequeña muestra de resultados. La

clasificación de las fuentes también trae consigo una importancia social sobre algún asunto o tema, y algunas fuentes pueden crecer o disminuir con el tiempo e indicar con ello un cambio en el compromiso y el nivel de atracción. Al leer los resultados de Google, podemos ver información e incluso algo sobre los propios mecanismos internos y autorales de Google (que abarcan resultados optimizados y manipulados), o bien, condiciones sociales (véanse la figura 0.1 y el capítulo 5 sobre la búsqueda como investigación). Este libro se enfoca en esto último.

Podemos hacer una exégesis similar en las redes sociales como Facebook y situar los métodos digitales una segunda vez. En este caso, es necesario resaltar no las diferencias entre el uso diario de un dispositivo y la mirada experta que revisa los resultados, como lo hicimos con Google, sino contrastar dos posturas de investigación en la web; por ejemplo: un amigo reciente tiene, a su vez, muchos amigos que actualizan sus perfiles y publican novedades que incluyen intereses sobre películas, música, libros y programas de televisión. En esa plataforma pueden crearse grupos del parque de juegos, de la secundaria, de la universidad, de colegas, así como de otros intereses sociales, y tendrán niveles cuantificables y potencialmente nuevas formas de socialización que los hagan cambiar. Después de todo, ahora el software maneja, en parte, la vida social y eso es algo sobre lo que hay que reflexionar.

No obstante, si se piensa con el dispositivo y se examinan los objetos digitales disponibles para su combinación, el trabajo del investigador cambia. También se puede pensar de la mano de los creadores de dispositivos y de los contenedores que ofrecen para que los usuarios llenen sus perfiles. ¿Cómo rearmar los objetos (amigos y perfiles) y reutilizar los resultados del dispositivo (perfiles y actividades de amigos) para poder ofrecer indicadores y hacer descubrimientos sobre la cultura (política)? Se puede considerar la recopilación de perfiles en formas reveladoras. ¿Qué nos dicen los intereses colectivos de los amigos de Barack Obama sobre las guerras culturales en comparación con los de sus opositores políticos? ¿Las posturas políticas están alineadas con las divisiones de gustos y preferencias o acaso las diferencias son menos grandes cuando se observan a través de las preferencias del medio

FIGURA 0.1 TIPOS DE DERECHOS: LA NACIONALIDAD DE LOS TEMAS

SUECIA	FINLANDIA	ESTONIA	LETONIA
 derechos humanos	 derechos de la infancia	 derechos de la ciudadanía	 derechos de los animales
 derechos de los y las pacientes	 derecho de acceso público a la naturaleza (libertad para deambular)	 derechos de la infancia	 derechos humanos
 derechos de la infancia	 derechos de los animales	 derechos ambientales	 derechos de personas usuarias del transporte aéreo
 derechos de personas usuarias del transporte aéreo	 derechos de las personas consumidoras	 derechos de personas usuarias del transporte aéreo	 derechos de pensión para personas no ciudadanas
 derechos de las personas creadoras	 derechos de las mujeres	 derechos de autoría	 derechos de los y las inmigrantes
 equidad de derechos	 derechos de personas usuarias del transporte aéreo	 derechos de los y las pacientes	 derechos de autor
 derechos de la ciudadanía	 derechos de las personas arrendatarias	 derechos de propiedad	 derechos de la infancia
 derechos de las mujeres	 derechos de los y las pacientes	 derechos de los y las terratenientes	 derechos sociales
 derecho a la negociación colectiva	 derechos de la juventud	 derechos de los y las trabajadores/as	 derechos del profesorado
 derechos alimentarios	 derecho a la educación en lengua de señas nativa	 derechos sexuales y de salud	 derechos de las personas consumidoras

Los principales diez derechos por país de acuerdo con dominios locales de Google, julio 2009. © Digital Methods Initiative, Amsterdam, 2009.

REINO UNIDO	PAÍSES BAJOS	BÉLGICA (flamenco)	BÉLGICA (francés)
 derechos humanos	 derechos de comité de empresa	 derechos humanos	 derechos humanos
 derechos de autoría	 derechos de personas usuarias del transporte aéreo	 derechos de personas con discapacidad	 derechos de internet
 derechos digitales	 derechos de la infancia	 derechos de los y las ciclistas	 derechos de la juventud
 derechos de las minorías	 derechos humanos	 derechos de las personas voluntarias	 derechos de la ciudadanía
 derechos de la ciudadanía	 derechos de las minorías	 derechos de personas usuarias del transporte aéreo	 derechos de propiedad intelectual
 derechos de empleo	 derechos de la población trabajadora sexual	 derechos de comité de empresa	 derechos de los y las pacientes
 derechos de publicidad	 derechos de los y las contribuyentes	 derechos de la infancia	 derechos de las mujeres
 derecho al aborto	 derechos de la juventud	 derechos de los y las solicitantes de empleo	 derechos de la infancia
 derechos de las personas que toman fotos	 el islam y los derechos de las mujeres	 derechos de los y las inmigrantes	 derechos de los y los trabajadores/as
 derechos de la infancia	 derechos de autoría	 derechos de los y las pacientes	 derecho a la autodefensa legal

(definidas de manera muy amplia)? ¿Las redes sociales son sitios para el estudio de los gustos compartidos?

Planteado de otro modo, este es un libro sobre investigación en internet que no se trata solo de internet. De acuerdo con el movimiento generalizado hacia el estudio de datos en la web (como mencionamos en las conclusiones), este libro busca ofrecer un objetivo para la investigación en internet que todavía no se ha hecho explícito: el desarrollo de una mirada y una concepción metodológica para la realización de investigaciones sociales en la web. En otras palabras, pretende que la investigación sobre internet vaya más allá del estudio de la cultura en línea y del estudio de los usuarios de las tecnologías de la información y la comunicación (TIC). En los siguientes capítulos presentamos diversos métodos digitales para trabajar con partículas diminutas (hipervínculos) y con enormes masas (redes sociales). En realidad, este libro puede ser leído como una historia de la investigación relacionada con internet, ya que ha evolucionado del hipervínculo, el análisis de sitios individuales y la creación de directorios a mediados de los años noventa (capítulos 2 y 3) hacia el estudio de los motores de búsqueda y la blogósfera de principios a mediados de la década de los 2000 (capítulos 4 y 5), hasta el crecimiento de las funciones de autolocalización, así como de la llamada Web 2.0 y las redes sociales a finales de los años 2000 (capítulos 6 y 7). Los capítulos reflejan el modo en que cada uno de estos conceptos se estudia con frecuencia, así como otras formas de hacerlo una vez que se emplean los principios de los métodos digitales.

Métodos digitales también aspira a plantear ciertos problemas que pueden ser abordados en la investigación de la web. El cuarto principio de los métodos digitales incluye la dificultad y los retos de usar datos de la web para hacer investigación social, ya que reformula el tema de la ubicación del punto de referencia. ¿En qué deben basarse los descubrimientos principalmente? De manera más específica, ¿los descubrimientos deben basarse en la red? ¿Acaso es necesario calibrarlos o compararlos con conjuntos de datos o lugares de estudio tradicionales (*offline*)? Este tema se puede presentar comparando dos proyectos: Google Flu Trends y un mapa de preguntas sobre recetas para el día de Acción de Gracias de los usuarios de allrecipe.com.

Ambos son proyectos sobre métodos digitales, pero funcionan con dos ideas muy distintas del punto de partida.

Google Flu Trends (desde 2007/2008) es un caso típico y didáctico de una forma de pensar a través de la disponibilidad de objetos nativamente digitales (indagaciones en motores de búsqueda y los lugares desde donde se busca), y la reutilización de resultados de los motores para hacer investigación social (los lugares de incidencia de la gripe).¹ El lugar de las indagatorias se usa para localizar en un mapa los brotes de gripe. A continuación, los resultados se comparan con los datos sobre ubicaciones de gripe en los centros de control de enfermedades de Estados Unidos, y los equivalentes nacionales y regionales de esta agencia gubernamental en los casi veinte países en los que Google monitorea esos brotes. El muestreo de los hechos encontrados en internet se compara con los datos de las agencias. Esta es la forma tradicional de muestreo en la que las condiciones de prueba no se buscan en la red a través de la reutilización de los métodos que ofrece el medio, sino offline, lo cual convierte a la web en un medio anticipatorio cuyas tendencias pueden ser confirmadas más adelante en otro lugar.

En contraste, podemos considerar otro proyecto que ha hecho muestreo en el mundo en línea, o al menos no lo ha hecho offline. Como lo reportó el *New York Times* con infografías o visualizaciones de datos que son, con frecuencia, el resultado de ciertos proyectos de metodología digital, las indagaciones en allrecipe.com un día antes de Acción de Gracias (el tradicional banquete estadounidense) se capturaron y se trazaron en un mapa geográfico.² En él se señalaron los lugares donde se buscaron recetas y, al hacerlo, se obtuvo una distribución geográfica de gustos o recetas preferidas a lo largo y ancho de Estados Unidos. Mientras que, durante años, las empresas de motores de búsqueda publicaron las principales búsquedas mensuales y anuales, las categorizaban ocasionalmente de acuerdo con los temas principales (ej. búsquedas políticas) y les asignaban nombre con fines de identificación de tendencias y mercadotecnia,

1. Ginsberg et al., 2008.

2. Severson, 2009.

como *buzz* y *zeitgeist*, los mapas de búsquedas de recetas agregan a la labor de los investigadores de resultados en los motores no solo la ubicación de las búsquedas, sino también el punto de vista de la investigación social. Muestran qué tipo de comida les gusta a las personas en ciertos lugares. Aquí la pregunta sería si a continuación el investigador se dirigiría a fuentes offline (encuestas telefónicas o quizá datos de venta de supermercados) o si seguiría trabajando con los datos obtenidos en la red, haciendo muestreos de los hallazgos hechos ahí mismo. ¿Se puede hacer un muestreo de los hallazgos obtenidos con indagaciones en motores de búsqueda a través de una investigación adicional de datos de la web, como, por ejemplo, fotografías de comida geolocalizadas? Los métodos digitales no necesariamente buscan hacer un muestreo de (todos) los hallazgos en línea, sino más bien cuestionarse si el estatus de la web tiene potencial como sitio de muestreo.

Estas primeras acciones y principios apoyados en métodos digitales entran en juego en este libro: seguir los métodos del medio conforme evolucionan, aprender cómo los dispositivos dominantes tratan a los objetos digitalmente nativos y pensar de acuerdo con esos tratamientos y dispositivos para recombinarlos o construir sobre ellos. La aspiración es reutilizar los métodos del medio para una investigación que no sea principal o únicamente sobre la cultura en línea. Los hipervínculos se convierten en medios no solo para evaluar el valor de un sitio web y asignarle una clasificación, que es como los tratan los dispositivos dominantes, sino también para mostrar políticas de asociación: la falta de reconocimiento (a través de la ausencia de enlaces) de las organizaciones no gubernamentales (ONG) armenias por parte de las instituciones intergubernamentales, por usar un ejemplo de la interpretación de un mapa de hipervínculos.

El análisis de vínculos también puede usarse con la finalidad de encontrar sitios relacionados y construir listas de URL. Dado que en Irán existe una lista de sitios censurados, ¿cómo pueden analizarse los hipervínculos para localizar sitios relacionados que no están en el registro de sitios que se está estudiando? El mapa señala los resultados de este muestreo dinámico, como yo lo llamo, y despliega los sitios

originales, así como aquellos localizados en forma reciente.³ Estos últimos, junto con los de la lista original, se alcanzan a través de proxys en Irán para poder determinar si cada sitio está bloqueado. El resultado es un mapa que muestra qué sitios están bloqueados y cuáles son accesibles, y después los indicadores del mapa señalan los bloqueados más nuevos (véase la figura 0.2). De hecho, es posible aprender más sobre la extensión de la censura usando análisis de vínculos (o muestreo dinámico de URL) para alargar la lista de URL más allá de la postura tradicional editorial, creadora de listas, como se describe en el capítulo 3 sobre análisis de sitios web.

Los sitios web archivados tradicionalmente se consideran una solución al error “404 file not found”. La llamada Wayback Machine del proyecto Internet Archive permite al usuario buscar y recuperar páginas de algún sitio web que ya no está activo, es inaccesible o ha sido editado. Las diversas páginas históricas de un sitio web archivado también pueden organizarse y presentarse de forma que cuenten alguna historia sobre la web —como el auge del algoritmo por encima del editor (humano), como veremos más adelante—. Los sitios archivados también se pueden recopilar, leer e indagar para examinar cambios en el tono y las opiniones de grupos sociales como la derecha política y su relación con el extremismo. ¿La derecha se ha vuelto más extremista con el tiempo mediante el uso de un lenguaje específico? Aquí se construye una colección de sitios del pasado hasta el presente y se comparan los cambios en el uso del lenguaje que en ellos aparece a lo largo del tiempo.

Los motores de búsqueda crean nuevos órdenes de las cosas en el mismo sentido en que clasifican las fuentes para cualquier tipo de temas. La lectura e interpretación de los resultados de los motores como jerarquías de fuentes verosímiles por tema puede ser, en sí misma, una forma de investigación social. La relación entre búsqueda e investigación, en ocasiones, se lamenta en las discusiones pedagógicas y la principal preocupación es que la búsqueda está tomando el lugar de la investigación en general, ya que los estudiantes googlean en vez

3. Rogers, 2009.

FIGURA 0.2 UNA RED DE CENSURA: SITIOS WEB SOCIALES, POLÍTICOS Y RELIGIOSOS IRANÍES



Sitios bloqueados recién descubiertos.

- <http://home.bip.net/radiohambastegi/>
- <http://ir.mondediplo.com/>
- <http://www.alahwaz.com/>
- <http://www.ashoob.com/>
- <http://www.asre-nou.net/>
- <http://www.bbc.co.uk/persian/>
- <http://www.channelonnetv.com/>
- <http://www.didgah.com/>
- <http://www.didgah.net/>
- <http://www.fadaee.org/>
- <http://www.fedayi.org/>
- <http://www.fwhi.org/>
- <http://www.geocities.com/-fedaian/>
- <http://www.gozareshgar.com/>
- <http://www.iran-jommelli.com/>
- <http://www.iranazad.com/>
- <http://www.irandokht.com/>
- <http://www.komalah.org/>
- <http://www.kvinnonet.org/>
- <http://www.mani-poesie.de/>
- <http://www.medusa2000.com/>
- <http://www.mihan.net/>
- <http://www.negahi.com/>
- <http://www.peykeiran.com/>
- <http://www.radio-international.org/>
- <http://www.radiobarabari.net/>
- <http://www.rdkl.com/>
- <http://www.roozonline.com/>
- <http://www.shahrvand.com/>
- <http://www.zagros.info/>

Método y hallazgos

Utilizando issuecrawler.net, rastrear direcciones URL (el 30 de marzo de 2006) desde la lista de la Iniciativa OpenNet de sitios iraníes sociales, políticos y religiosos. Encontrar redes de sitios a través de una iteración de análisis de coenlace (ver figura). Traer cada sitio a través de un proxy en Irán (el 18 de mayo de 2006). Descubrir sitios que no se sabía que estaban bloqueados. Comparar sitios bloqueados recién descubiertos con la lista completa de los sitios web iraníes revisada por la Iniciativa OpenNet. El hallazgo es el descubrimiento de 30 sitios bloqueados en Irán.

Visualizaciones de la censura de internet 2006

Una red censurada: Sitios sociales, políticos y religiosos iraníes



Datos, por la Iniciativa OpenNet y Mapa, por issuecrawler.net. Análisis, por Richard Rogers, Nart Villeneuve y Erik Borra; y diseño, por Marieke van Dijk.

© 2006 Govcom.org y ONI

Legenda

- Sitios bloqueados
- Sitios bloqueados recién descubiertos
- Sitios accesibles
- Sitio inaccesible (debido a un problema con el DNS)

El resultado de muestreo dinámico de URL, método de análisis de hipervínculos para el descubrimiento de sitios web censurados. Gráfico de Govcom.org, 2006. © Govcom.org Foundation, Amsterdam, 2006.

de visitar bibliotecas y usan Wikipedia como una primera fuente de facto más que cualquier otra que sea menos folclórica y más verificada por expertos.⁴

No obstante, los motores también pueden ser reutilizados a fin de llevar a cabo investigaciones gracias a su capacidad de indexar sitios web y de ofrecer formas de búsqueda en sitios individuales o en conjuntos temáticos. Entre los (principales) sitios web que tienen que ver con el cambio climático, ¿cuáles son los que mencionan a los escépticos y con qué frecuencia y fervor? ¿Acaso los escépticos empiezan a prevalecer en las principales fuentes sobre cambio climático? Aquí la web se convierte en un lugar de investigación al expandir los tipos de fuentes, por ejemplo, más allá de lo científico (y su metafuente ISI Web of Science) y las noticias (Lexus Nexus). De ahí surgen otras preguntas sobre el uso de la web como lugar de investigación. ¿Los motores están demarcando los conjuntos de fuentes e indexando fuentes individuales con la misma conciencia y rigor de sus contrapartes en la investigación tradicional? ¿Qué tan sólida es la demarcación de fuentes y los hallazgos en la web que de ella se desprenden? ¿En qué condiciones pueden sostenerse junto a aquellas derivadas de las técnicas de la cuantimetría y de la investigación de atención a la prensa? La comparación de los métodos digitales con las técnicas más tradicionales va más allá de la investigación de los medios. ¿El registro de datos de un motor de búsqueda puede competir en términos de los resultados de una investigación con las encuestas y las preferencias identificadas en redes sociales?⁵

Si bien este libro puede leerse como una historia de la investigación relacionada con la web, también se enriquece con la Digital Methods Initiative, un proyecto que inicié en 2007 como programa de investigación en la Universidad de Ámsterdam, un wiki (digitalmethods.com) y un curso de verano. La Digital Methods Initiative (DMI) recibió una subvención de la Mondriaan Foundation como parte del programa *interregeling*, dedicado a proyectos que no encajan en ninguna otra

4. Nyirubugara, 2011.

5. Lewis et al., 2008; Neff, 2011; Tumsajan et al., 2010; Gayo-Avello et al., 2011.

categoría. Se trata del campo denominado “otro” en los formularios (que ya ha sido discontinuado). El primer método digital que se enseñó en el entonces pequeño e informal curso de verano recibió el nombre de “distancia de la fuente”. ¿Qué tan lejos están las historias y las fuentes en las listas de los medios? Puede que una historia se encuentre al principio de la lista en los medios de comunicación, un titular. Puede ser una historia de primera plana en un periódico. ¿Cómo se determina la lista en la web? Un caso de estudio ejemplar tiene que ver con los escépticos del cambio climático. En un momento dado, la BBC News anunció en su noticiario la cancelación del programa Planet Relief, dedicado a la concientización y creado con base en el esquema de Live 8 (2006) y Live Earth (2007). Los motivos de la cancelación eran que la BBC no debía tomar partido, y que Planet Relief no debía ofrecer a los escépticos la oportunidad de expresar sus puntos de vista. Los activistas medioambientales citados en una nota periodística sobre la cancelación del programa dijeron lo siguiente:

La única razón por la que esto es un tema hoy día es que existe un pequeño pero vociferante grupo de “escépticos” climatológicos que cabildan contra las acciones concretas, de modo que la BBC actúa cobardemente y se rehúsa a adoptar una postura más consistente.⁶

En el caso de la nota sobre la cancelación de la BBC y otros temas sobre cambio climático, los escépticos aparecían de manera constante en las primeras planas. La pregunta que se planteó a los participantes del curso de verano de Digital Methods fue la siguiente: ¿los escépticos se encuentran también en los primeros lugares de la lista en la web? La “distancia entre las fuentes” mediría la distancia entre la parte alta de la web y las fuentes afines a los escépticos, de acuerdo con la herramienta dominante Google, como se explica con más detalle en el capítulo 5. En nuestros procedimientos de investigación, se consulta Google sobre el cambio climático y se guardan los resultados; enseguida, se busca en cada uno de los resultados a los escépticos del

6. Black, 2007: 1.

cambio climático y se anota su lugar en la clasificación de fuentes. Posteriormente, se calcula la distancia entre el lugar que ocupan y la parte más alta de la lista.⁷

La distancia entre las fuentes se convirtió en una técnica útil no solo para el análisis de la “esfera web” (un término creado por los investigadores de archivos en la red y que mis colegas y yo nos apropiamos para hablar de análisis comparativo de medios o, en realidad, la versión de la web de ahí surgida). Como punto de inicio, mostramos cómo el análisis de medios entre tipos de fuentes como periódicos, revistas y noticieros se ha llevado a cabo en el pasado e influyó en la realización de la misma actividad en la web, comparando las esferas dominantes en el mundo de internet de entonces: esfera web, blogósfera y esfera noticiosa. Las esferas se interpretan como espacios demarcados por motores. Cada una tiene un motor dominante (principalmente Google, aunque por un tiempo Technorati fue un motor de búsqueda dominante en la blogósfera), y cada motor tiene una lógica general distinta para clasificar las fuentes en cada esfera. La clasificación para la esfera web es diferente a la de la blogósfera y de la esfera noticiosa. Por tanto, las esferas invitan a la comparación. En virtud de ciertos temas o historias, ¿cuáles son las fuentes en la parte alta de sus respectivas esferas? El “análisis interesférico” se trata de la comparación de la distancia entre las fuentes a través de la esfera web, la blogósfera y la esfera noticiosa (más adelante agregaríamos otras esferas).

La distancia entre las fuentes se presta para el análisis de la blogósfera dado que permite tanto la crítica como el análisis empírico de la “esfera”. Para algunos, la idea de “esfera” en la blogósfera invita a pensar en la esfera pública. La igualdad de las voces, el ideal igualitario, también es evidente en la forma geométrica de la esfera en el que las fuentes son equidistantes al centro. De hecho, la esfera en la blogósfera inicialmente amenazaría la jerarquía de las listas clasificadas (y

7. De acuerdo con las convenciones que utiliza Google, las consultas en los motores de búsqueda se colocan entre corchetes en esta discusión, como [9/11]. Si se usan comillas en una consulta (para obtener resultados exactamente iguales), la consulta se registra de esa forma: [“cambio climático”]. Cuando se dejan fuera las comillas, los resultados de una consulta son equivalentes o sinónimos: por ejemplo, la consulta [teléfono celular] también entrega resultados como teléfono móvil.

los resultados de los motores de búsqueda) que alguna vez pusieron orden en la web antes del crecimiento de la blogósfera. Con esta nueva forma, la blogósfera rompería la jerarquía de las fuentes (incluso si en la práctica existía una referencia continua a los bloggers de primera categoría usando el lenguaje de la cultura de las noticias del entretenimiento). Si todos los blogs estaban en una posición equidistante en relación con el centro, entonces, en principio, todo blog tenía posibilidades de ser conocido por todos.

Estos espacios y geometrías infopolíticas (listas clasificadas, redes, esferas, la llamada *dark web* o web oscura) son el tema del capítulo 2 sobre hipervínculos y las políticas del espacio en la web. Este capítulo se escribió originalmente para *The Handbook of Internet Politics* (2007) y han aparecido otras versiones en volúmenes de *Nouvelles technologies cognitives et épistémologie* (2007) y *Digital Cognitive Technologies: Epistemology and Knowledge Society* (2010), ambos editados por Claire Brossard y Bernard Reber. Una versión un poco más larga se publicó en *Theory, Culture and Society* (2012), en el fascículo especial sobre la aproximación topológica a la dinámica cultural, editado por Celia Lury.

El análisis de los motores de búsqueda como “inocentes”, tema del capítulo 4, es en realidad una reacción a los medios a través de los cuales la personalización de los resultados afectaría la investigación sobre la distancia entre fuentes. En la investigación sobre la epistemología de la web, tanto la nueva jerarquía de las fuentes como la credibilidad de los resultados de los motores resultan de interés. Si no existen dos individuos que reciban los mismos resultados a la misma búsqueda (desde diciembre de 2009), entonces resulta inapropiado “culpar” a los motores de búsqueda por colocar a los escépticos del cambio climático (por seguir usándolos como ejemplo) en la parte superior de la lista de la web cuando se hacen búsquedas sobre cambio climático.⁸

Con la personalización, los resultados de los motores de búsqueda son de la autoría tanto del motor como del usuario; es decir, los resultados que uno obtiene son, en parte, consecuencia de nuestras propias acciones con base en el historial de búsquedas, la ubicación

8. Pariser, 2011.

y otros marcadores, tal como Google se refiere a los puntos de datos que ha ido recolectando. El tema entonces es si uno puede entrenar a un motor de búsqueda (cuenta) a ser “amigable” con los escépticos y a colocar las fuentes que los mencionan en la parte alta de la lista de resultados cuando se hace una búsqueda [“cambio climático”], y a otro motor de búsqueda (cuenta) a no ser amigable con los escépticos y colocar las fuentes que los citan en la parte inferior de la lista. De manera similar, podemos imaginar el deseo de entrenar a un motor de búsqueda a tener cierta postura política, y otra para algún otro motor.

Este tipo de trabajo que llevan a cabo algunos colegas abre de par en par el estudio del impacto de la personalización en los resultados de los motores, un importante tema de la epistemología de la web.⁹ (Yo mismo he entrenado a un motor de búsqueda a ser propleles animales y a otro a ser antipleles animales.) El artículo sobre el “motor inocente” se publicó en *Deep Search: The Politics of Search beyond Google* (2010) junto con la versión alemana *Deep Search: Politil des Suchens jenseits von Google*, editadas por Konrad Becker y Felix Stalder. También se trata de una contribución propia para académicos de la *googlización*, en particular Siva Vaidhyanathan, como parte de un esfuerzo por dar forma a la agenda de investigación.¹⁰

Los cursos de verano de Digital Methods (2007-) y los talleres ofrecidos por Govcom.org (1998-) generaron mucho del material que se usa en este libro, incluidas las contribuciones que hacemos al estudio del sitio web y de las redes sociales. Los sitios web son predominantemente azules, como lo descubrimos en el taller *Recalling RFID*, organizado en Balie, Ámsterdam, en 2007 (véase la figura 0.3). Más allá del análisis del color y la usabilidad, un sitio web puede estudiarse de acuerdo con su genealogía (plantilla), anatomía, características y otros puntos de partida que se mencionan en el capítulo 3.

Un dispositivo que organiza sitios web como objetos de estudio es la Wayback Machine de Internet Archive. Con frecuencia, las in-

9. Feuz et al., 2011

10. Vaidhyanathan, 2011.

FIGURA 0.3 ESTUDIO DE LOS COLORES DE LOS SITIOS WEB EN EL ESPACIO DE EMISIÓN DE RFID, EN ALUSIÓN AL TALLER RFID, ÁMSTERDAM, 2009

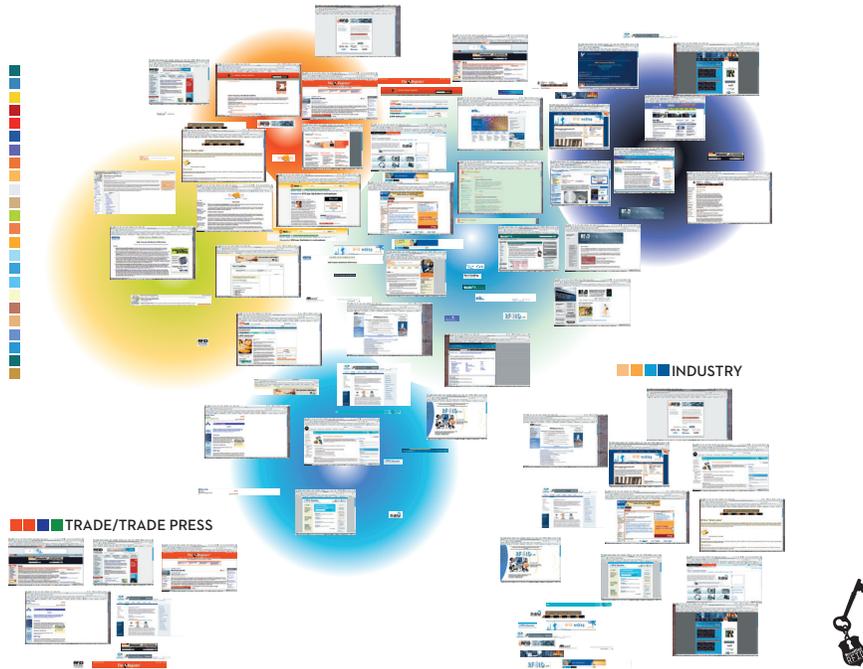


Gráfico de Studio Leon&Loes, Rotterdam. © Studio Léon&Loes, 2009.

investigaciones lamentan la falta de uso de esta herramienta, así como los archivos nacionales web de otros académicos e investigadores, a excepción de equipos legales que buscan evidencia o departamentos legales que indagan en sus propios sitios web y que piden que se eliminen o que se inserte user-agent: ia_archiver Disallow: /en el archivo robots.txt, lo cual no solo excluye los rastreadores o indexadores web, sino que también borra el historial del sitio del archivo.

¿De qué otra forma se puede estudiar el sitio web como objeto archivado? En el curso de verano de Digital Methods (que fue también el décimo aniversario de Govcom.org) aplicamos los principios de los métodos digitales siguiendo el dispositivo dominante

y reutilizándolo con fines de investigación. En la Wayback Machine se ingresa un URL que recupera las páginas archivadas. En esencia, se organiza la historia de la web en las historias de un solo sitio o una sola página. Tiene integrada una historiografía inmersa en la tradición biográfica. Al considerar cómo reutilizar este resultado, mis colegas y yo capturamos las páginas históricas de un URL (Google.com, también durante su décimo aniversario). Recopilamos las páginas singulares (las que tenían un * al costado en la página de resultados de la Wayback Machine), las cargamos en una presentación que se reprodujo al estilo de la fotografía secuencial o de los documentales que utilizan la técnica conocida como *screencast* (es decir, grabar aquello que ocurre en la pantalla de la computadora) con una voz en *off* que cuenta la historia de Google a través de los cambios en su interfaz desde 1998 hasta 2007, como parte de un proyecto llamado *Google y la política de las “pestañas”/Google and the Politics of Tabs*. (El archivo de internet normalmente tiene seis meses de retraso o más en la lista de contenido de sitios web archivados.)

En 2009 se echó a andar un proyecto más ambicioso como parte de los cursos de verano de Digital Methods. ¿Cómo avanzar más allá de las historias de un solo sitio web? El objetivo inalcanzable que se fijó fue el de conjurar un estadio anterior de la web para permitir el estudio de un periodo particular en vez de solo revisar la biografía de un sitio único como hicimos en el caso de la película de Google. Además de las biografías de sitios, existen historias de eventos organizadas en muchas de las colecciones especiales de ciertos sitios, por ejemplo, la colección del 11 de septiembre de 2001 o las dedicadas a desastres naturales o ciertos procesos electorales, como comentaremos más adelante.

Conforme a lo que podemos llamar una perspectiva de plataformas de nuevos medios más general (no una herramienta, sino un creador de herramientas) y con esa misma práctica de archivado de la web contemporánea (el proyecto Archive-IT), mis colegas y yo presentamos una técnica para crear una colección de sitios ya archivados para estudiar la historia de la web, o la historia con ella. Lo que queda de la blogósfera temprana se determinó y se capturó para llevar a cabo lo que parece que es uno de los primeros análisis o mapeos de

hipervínculos. En cualquier caso, los primeros que no están archivados cobran vida en el mapa de hipervínculos al mostrar no solo su presencia y posicionamiento a través de los vínculos que tuvieron en su momento (agosto de 2000), sino también un estadio pasado de (parte de) la web, incluidas las esferas relativas de la blogósfera. La técnica de conjurar sus estadios pasados se ha aplicado en forma subsecuente a la blogósfera holandesa, así como a la web palestina; es, por tanto, un método digital (como la distancia entre las fuentes y otras más) que ha sido adoptado de manera amplia.

La continua relevancia del trabajo de la etapa temprana de los cursos de verano de Digital Methods (2007, 2008 y 2009) se hizo clara por los métodos que han perdurado hasta hoy y por la distancia entre fuentes que se incluye en el Dispositivo Lippmanianno (que comentamos en el capítulo 5), y los documentales que usan la técnica screencast, además de los análisis históricos de vínculos de páginas en los archivos de la web (capítulo 3). Ciertos proyectos incipientes de los primeros días maduraron y se convirtieron en esfuerzos para demarcar y diagnosticar la condición de la web iraquí, en el verano de 2007, pasados cinco años del inicio de la guerra en Iraq, cuando las voces de los blogueros en tierra (Salam Pax y el llamado Baghdad blogger) y de los senadores que hacían viajes para conseguir información iban al mercado de Baghdad (cámara en mano) en un intento por brindar narraciones fidedignas de las condiciones locales. ¿Era posible agregar aquellas narraciones a la situación en Iraq analizando el estado de salud de su web? Al hacer la recolección de sus sitios web, encontramos una web rota, en la que los sitios de la universidad no recibían mantenimiento o estaban del todo caídos, por ejemplo. Uno de los pocos activos era el del Ministerio Iraquí del Petróleo, iel único en toda la colección que tenía anuncios! Durante el proyecto, insistimos en hacer una aproximación al estudio de la web nacional que se basara en una serie de mediciones para diagnosticar su estado de salud, como en el caso de la capacidad de respuesta y frescura de las páginas web nacionales, que se aplican en el capítulo 6. El caso de estudio es Irán, en donde, en 2011, demarcamos una web nacional (más de diez mil servidores individuales) basándonos en las llamadas “culturas de dispositivos” dominantes entre los iraníes, es decir, motores y plataformas que repetidamente obtienen

datos de los usuarios y les ofrecen sugerencias de URL. Al estudiarlo, no obstante, surgió una complicación adicional: muchos de sus sitios web estaban censurados por el Estado. En una investigación llevada a cabo por el Programa Iraní de Medios (Centro de Estudios Globales en Comunicación, la Escuela Annenberg de Comunicación, la Universidad de Pensilvania), Esther Weltevrede, Sabine Niederer, Erik Borra y yo nos encontramos con una web censurada, pero viva, en la que los blogueros escribían a pesar de que los sitios estaban bloqueados. También presentamos una serie de mediciones para estudiar la salud de la web nacional, al estilo de un índice de la web. Una versión más corta de la investigación apareció en *Blackwell Companion to New Media Dynamics*, editado por John Hartley, Jean Burgess y Axel Bruns.

Las ideas desarrolladas para el estudio de las redes sociales, así como el término “posdemográfico”, tema del capítulo 7, surgieron en “Space for People: Suggested Fields”, un taller apoyado por el Instituto de Arte de Nuevos Medios de los Países Bajos y Montevideo/Time Based Arts, y en el evento “Walled Garden”, organizado por Virtual Platform, ambos en Ámsterdam en 2008. Los resultados fueron proyectos de software de arte de nuevos medios, como *Elfriendo.com* y *Leakygarden.net*, aunque solo el primero fue objeto de cobertura mediática por parte de una revista de arte (para jóvenes) en Ámsterdam.¹¹ Es importante señalar que ambos proyectos tenían como punto de partida ideas compatibles con dispositivos en línea y la construcción sobre estos con fines de investigación en redes sociales y, más específicamente, de los perfiles en línea.

Elfriendo, que tenía el eslogan “sin las complicaciones de las redes sociales”, sería capaz de crear un perfil personal desde cero basándose en el ingreso de un solo tema de interés. Existían otros dos servicios dentro de *Elfriendo.com*: se revisaba la compatibilidad entre dos perfiles y se ofrecía una “mejora” de perfil. Construido encima de Myspace, y en particular de la búsqueda de intereses en Myspace (que sería descontinuada más adelante), el software buscaba entre los perfiles (de la parte alta de la lista) de quienes tenían un interés particular o un tipo

11. El proyecto artístico *Elfriendo.com* apareció en *Mister Motley*. Véase Toma, 2008.

de música, películas, programas de televisión, libros o héroes favoritos.¹² Los recopilaba y creaba con ello un nuevo perfil o la renovación de este (los campos de cada usuario se “refrescaban”, por decirlo de alguna manera). También comparaba dos perfiles y ofrecía porcentajes de compatibilidad. Entre las cosas importantes, contrastaba los conjuntos de perfiles con intereses particulares. ¿Cuál es la suma del perfil de los interesados por Barack Obama y de los interesados en John McCain, oponente de Obama en la elección presidencial de 2008? ¿Son compatibles? ¿A qué grado los interesados en Obama escuchan la misma música, ven las mismas películas y programas de televisión, leen los mismos libros y tienen los mismos héroes que los interesados en otros candidatos?

El término “posdemografía” invita a estudiar de manera social o cultural perfiles significativos en su conjunto y a indagar la relación entre ellos. También es una invitación a preguntar sobre lo que tienen en común. ¿El islam es compatible con el cristianismo en el sentido de que quienes lo incluyen como un interés comparten otras cosas favoritas? El capítulo sobre redes sociales se escribió originalmente como “Post-demographic Machines” (Máquinas posdemográficas) en *Walled Garden* (2009), editado por Annet Dekker y Annette Wolfsberger, y una versión ampliada apareció como “Post-demographic onderzoek: voorbij het doelgroepen denken” (Investigación posdemográfica: más allá del pensamiento basado en el grupo meta) en el fascículo especial de *De Gids* (2010), que exploraba la reescritura en la web del canon literario neerlandés, editado por Noortje Marres.

12. Los URL (descontinuados) para los dispositivos de búsqueda de Myspace y sus breves explicaciones, son los siguientes: [http://searchservice.myspace.com/index.cfm?fuseaction=searchandsearchtarget=tpeopleandsearchType=networkandsearchBoxID=Profileandinteresttype=Gandcountry=andsearchBy=FirstandSubmit=Findandf_rst_name=andf_search_criteria](http://searchservice.myspace.com/index.cfm?fuseaction=searchandsearchtarget=tpeopleandsearchType=networkandsearchBoxID=Profileandinteresttype=Gandcountry=andsearchBy=FirstandSubmit=Findandf_rst_name=andf_search_criteria;); http://searchservice.myspace.com/index.cfm?fuseaction=searchandsearchtarget=tpeopleandsearchType=networkandsearchBoxID=Profileandinteresttype=Muandcountry=andsearchBy=FirstandSubmit=Findandf_rst_name=andf_search_criteria; http://searchservice.myspace.com/index.cfm?fuseaction=searchandsearchtarget=tpeopleandsearchType=networkandsearchBoxID=Profileandinteresttype=Mandcountry=andsearchBy=FirstandSubmit=Findandf_rst_name=andf_search_criteria; y http://searchservice.myspace.com/index.cfm?fuseaction=searchandsearchtarget=tpeopleandsearchType=networkandsearchBoxID=Profileandinteresttype=Bandcountry=andsearchBy=FirstandSubmit=Findandf_rst_name=andf_search_criteria. Interesttype= Aquí, G, Mu, M, y B son intereses generales, música, películas y libros, y f_search_criteria= puede llenarse con los intereses específicos que se buscan.

This Painting Is not Available in your Country, pieza artística de Paul Mutant (2010), que resume la idea del fin de lo virtual, tema del capítulo 1, se trata de la conferencia que dicté durante la inauguración de la cátedra sobre nuevos medios y cultura digital en la Universidad de Ámsterdam. Fue publicado como “The end of the virtual: Digital methods” (“El fin de lo virtual: Métodos digitales”) (2009) por Amsterdam University Press, y en versión resumida en alemán en el *Zeitschrift für Medienwissenschaft* (2011).

Tanto en “El final de lo virtual” como en “La política en el espacio de la web” (ahora el capítulo 2) destaca la importancia del advenimiento de la tecnología *IP-to-geo* como la venganza de la geografía en el ciberespacio. La fórmula más dramática es “la muerte del ciberespacio”, que se refiere a que la web es cada vez menos un lugar —si es que lo fue alguna vez—, tomando en cuenta la cercana relación entre la comunidad virtual y los lugares reales que menciona la literatura de principios y mediados de la década de los noventa, y cómo lo local se apropia de internet para cumplir con sus propios propósitos, como lo descubrieron los etnógrafos al visitar cibercafés en Trinidad y Tobago en los años 1990; es decir, la tecnología de geolocalización ubica al usuario y carga (o lo bloquea) contenido a la medida de acuerdo con el buscador del usuario. Esta tecnología se implementó ampliamente como resultado de una demanda contra Yahoo! en Francia en 2000 por parte de ONG dedicadas a la antidifamación, y se aplicó con rapidez en contextos publicitarios y de propiedad intelectual, como en el *streaming* de programas de televisión, incluidos los Juegos Olímpicos de 2008, en el que muchos experimentaron sus efectos, como fue posible observar. Los medios de localización eran normales.

Los dominios locales de Google (Google.fr, por ejemplo) también son ejemplos de los efectos de la geolocalización, ya que el Google que se carga por defecto es el que está asociado a la ubicación del usuario. En la mayoría de los dominios locales de Google hay una liga que lleva a “Google.com en inglés”, y también hay un URL separado (Google.com/ncr) que se ofrece para que no se “redirija” a algún país, lo cual suprime la autodetección de la ubicación. La falta de lugar o ubicación, que alguna vez fue la experiencia y la retórica del ciberespacio, es hoy un URL alternativo.

El capítulo 8, sobre Wikipedia como referencia cultural, es una colaboración con Emina Sendijarevic, quien en el curso Digital Methods, entre 2010 y 2011, en la Universidad de Ámsterdam estudió con mucho cuidado los artículos sobre la caída, la masacre y el genocidio de Srebrenica en las páginas de Wikipedia en neerlandés, serbio y bosnio, en ese orden (así como otras más). Con algunas excepciones (el artículo neerlandés, por ejemplo), los artículos tuvieron su origen a partir de una traducción del inglés; a lo largo de seis años de ediciones, se crearon narrativas y puntos de vista particulares que describimos, originalmente, como puntos de vista nacionales más que neutrales, tal como se le conoce a este principio de Wikipedia. En el análisis nos interesamos en saber si podríamos sustraernos del marco de Wikipedia como enciclopedia y navegar hacia un debate preciso. Proponemos que el estudio comparativo de artículos de Wikipedia entre versiones de diversos idiomas es una referencia cultural, más que estándar o universal, lo cual se ciñe al principio de los métodos digitales del uso de la web para estudiar los cambios, preferencias y compromisos culturales.

Inmediatamente después de que casi la misma versión del artículo se subió a las Wikipedias en bosnio, croata y serbio en 2005, un usuario serbio subió la plantilla para disputar el punto de vista neutral, mientras que los usuarios bosnios y croatas elogiaron su calidad y, por último, convirtieron la versión en bosnio en un artículo relevante, una especie de medalla de honor que se da a los artículos notables. Tras cinco años de edición, los artículos en neerlandés y serbio tienen un conteo menor de víctimas que el artículo bosnio (y que el resto de los artículos analizados), y ponen de relieve las muy diversas fuentes (y opiniones) de estos hechos. Si bien Srebrenica es un caso especial, el método con el que se estudia el “mismo” artículo en diferentes idiomas en Wikipedia es un hecho generalizable; será el tiempo el que diga si el método se adopta.

“Después del ciberespacio: macrodatos y microdatos”, el capítulo final, ubica los métodos digitales una vez más al volante computacional de las ciencias sociales y las humanidades digitales, y busca hacer distinciones entre una serie de aproximaciones contemporáneas al estudio de los datos digitales y el que presentamos aquí. Al principio

del último capítulo, así como en la introducción, volvemos a la diferencia entre lo digitalmente nativo y lo digitalizado, tanto en forma de datos como de método. Ciertas partes del capítulo se presentaron en el congreso MiT17 sobre Plataformas Inestables en 2011, y también toma prestadas algunas ideas que escribí en el contexto de dos proyectos con recursos de la Comisión Europea: *A Topological Approach to Cultural Dynamics* (Aproximación topológica de la dinámica cultural), el sexto proyecto marco ya mencionado, y *Mapping Controversies on Science for Politics* (Controversias de mapeo en la ciencia de lo político, Macospol, por sus siglas en inglés), el séptimo proyecto marco encabezado por Bruno Latour en Sciences Po, París.

El proyecto Macospol también tuvo como resultado la transformación de Google Scraper (una herramienta de investigación de distancia entre fuentes) hasta convertirla en el llamado Dispositivo Lippmaniano (la herramienta que se usa para analizar preferencias partidistas y compromiso con ciertos temas). El grupo de Digital Methods en Ámsterdam creó la plataforma de mapeo de controversias de Macospol en mappingcontroversies.net.

La distinción entre lo digitalmente nativo y lo digitalizado se hace como un cambio de postura.¹³ Muchos de los acercamientos metodológicos al estudio de los nuevos medios digitales trabajan con datos digitalizados, como el “análisis cultural” y la llamada “culturomics”, como lo mencionamos en la conclusión del capítulo. Uno busca patrones entre conjuntos completos de pinturas de un mismo autor o las portadas de revistas de estilo de vida, mientras que el otro estudia los cambios en el uso de las palabras a lo largo del tiempo buscando en libros escaneados por Google (como parte del proyecto Google Books). En otras palabras, en el caso de estas y otras aproximaciones, los medios digitales se refieren al estudio de lo digitalizado y lo escaneado, en el que, con frecuencia, se requiere acceso a los datos y privilegios de búsqueda.

En contraste, aquí presentamos una estrategia para usar las formas y materiales de medios digitales específicos (como las publicaciones

13. El término “nativo” se usa en el sentido computacional. Una aplicación es nativa cuando se ha escrito especialmente para un sistema específico.

en blogs y la edición de Wikipedia) en vez de usar datos que han sido digitalizados o escaneados y cargados en un medio digital (como en las fotografías de la Segunda Guerra Mundial que se suben a una página web, que señalamos en el capítulo 3 como ejemplo de la selección que hace un archivista web de sitios valiosos que deben ser archivados). El trabajo de este libro sobre el sitio web como objeto archivado se escribió en la primavera de 2011, cuando fui miembro de la Escuela Annenberg de Comunicación de la Universidad de Pensilvania. Me gustaría agradecer a Joseph Capella, Joseph Turow y Michael Delli Carpini por llevarme a Filadelfia. Klaus Krippendorff inspiró una versión de análisis de contenido digital de métodos digitales en la investigación de Wikipedia que ya cité con anterioridad.

También se requiere acceso especial para el estudio de ciertos objetos digitalmente nativos como los *logs* o registros de búsqueda que ofrecen los buscadores, lo que lleva a la discusión sobre la tensión entre el uso de las aplicaciones de programación de interfaces o API que ofrecen las empresas y la extracción de datos. Es necesaria una revisión ética acompañada de propuestas para extraer datos en línea, incluidos los perfiles.¹⁴ En parte, se ha heredado de la decepción que originó, en 2006, la publicación de historiales de búsqueda de usuarios de AOL supuestamente anónimos, tras lo cual varios periodistas pudieron “desanonimizar” o identificar a uno de los usuarios del motor de búsqueda, una señora mayor en Georgia. Los datos de AOL se liberaron para que los científicos los usaran a fin de mejorar las búsquedas (personalizadas). Aquí hacemos un contraste entre la postura del *voyeur* o el detective que está aparejado con usuarios de motores de búsqueda numerados y con listas de sus búsquedas—los formatos de datos que se ofrecen—con otra postura, más en sintonía con las necesidades de investigación social. Para la postura de la investigación que pretendemos describir como métodos digitales, los datos elegidos no son los individuos, sino los lugares de búsquedas sustantivas.

Es de reconocer el trabajo de DMI en Ámsterdam: Anat Ben-David, Erik Borra, Marieke van Dijk, Anne Helmond, Koen Martens, Sabine

14. Zimmer, 2010a.