

Interpretación histórica y objetos digitales: consideraciones a partir de ejemplos concretos

*Historical Interpretation and Digital Objects:
Considerations based on Concrete Examples*

Nicolás Quiroga
Universidad Nacional de Mar del Plata
<https://orcid.org/0000-0002-1276-4476>
nfquirog@gmail.com

Recibido: 19/07/2021; Revisado: 21/12/2021; Aceptado: 29/12/2021

Resumen

Uno de los problemas fundamentales para la investigación histórica con objetos digitales se relaciona con su interpretación. A las problemáticas sobre la preservación de los objetos digitales ya consideradas, especialmente por la archivística, hay que agregar las cuestiones relacionadas con el análisis de colecciones digitales sin curaduría, semiestructuradas, parciales, etc. Partiendo de una conocida definición de objetos digitales que propone pensar sus capas física, lógica y conceptual (K. Thibodeu), el artículo discute las implicancias para el análisis histórico de esas consideraciones a partir de casos concretos, con colecciones de objetos digitales diversas en su extensión y propósitos: *logs* de chats y colecciones enormes como la de Geocities.com.

Palabras clave: Historia digital, objetos digitales, archivo.

Abstract

One of the fundamental problems for historical research that employs digital objects is how these latter should be interpreted. In addition to issues relating to the preservation of digital objects, already covered by archival science in the main, there are also questions around the analysis of digital collections without curation, or that are semi-structured, partial, etc. Departing from a well-known

definition of digital objects that proposes to think about their physical, logical and conceptual layers (K Thibodeu), this article discusses the implications of said definition for the historical analysis of digital objects on the basis of concrete cases, using collections of digital objects of differing volume and purpose: chat logs and huge collections such as Geocities.com.

Keywords: Digital History, Digital Objects, Archive.

1. INTRODUCCIÓN

Desde hace décadas, diferentes disciplinas abocadas al estudio del pasado se interrogan acerca de las transformaciones que implica la consolidación de un nuevo tipo de materialidad documental, un nuevo tipo de soporte asociado a las computadoras, una forma de inscripción capaz de volverse sustantivo y pasar a ser denominador de época: «lo digital». Pese a ser, como bien indica B. PETERS (2016), un fenómeno propio del siglo XX, la conversación sobre «lo digital» (entendido como «giro» o como «era») es inagotable. Si la enfocamos en su relación con la investigación histórica, el tópico atraviesa múltiples problemáticas, tales como materialidad del texto, archivo, interpretación, metodología, comunicación de la ciencia (WINTERS, 2020). Desde los años noventa del siglo XX, la digitalización masiva y la multiplicación de documentos nacidos digitales habilitaron una conversación que no dejó de interrogarse sobre las transformaciones de los medios y las mediaciones, pero aún más sobre los cambios que las tecnologías introducían o podían introducir en el oficio (PUTNAM, 2016; CAIMARI, 2017).

Pero esa conversación no se hizo urgente ni decisiva para la investigación histórica ni para la formación académica. Tal vez esta sea una evaluación apresurada de 30 años de discusiones de la *digital history*, pero algunas intervenciones recientes lo sugieren (CRYMBLE, 2021): los aportes conceptuales y metodológicos de la (angloparlante) historia digital no significan cambios paradigmáticos o definen sustanciales contribuciones al conocimiento histórico. En otras palabras, y como suele suceder con otros oficios, es posible continuar trabajando de manera «tradicional», bajo marcos de significación más próximos al *Dutch Manual* (incluso si no conocemos la relación de ese noble texto con el principio de procedencia) (BAILEY, 2013). Aun así, desde dos dimensiones del oficio podríamos relativizar ese último diagnóstico: el archivo no textual y los documentos nacidos digitales.

El objetivo de este artículo es informar sobre problemáticas conceptuales, metodológicas y prácticas surgidas en el trabajo con fuentes nacidas digitales. Voy a exponer dos escenarios concretos en los que tuve que investigar («hacer archivo») con objetos digitales. Con el primero reflexiono sobre el análisis de las sociabilidades en la prehistoria de la web a partir de registros de conversaciones en *Bulletin Board System* (BBS), y me interrogo sobre la interpretación de objetos digitales y su relación con la tradición interpretativa de la cultura escrita, a partir de preguntas tradicionales sobre la sociabilidad. A partir del segundo escenario me pregunto sobre cómo investigar colecciones inabarcables, enormes cantidades de datos sin curaduría a partir de los restos de *Geocities.com*. Al final del artículo arriesgaré unas pocas líneas sobre lo que podrían ser dos objetivos en siguientes

proyectos de investigación.

Antes de adentrarnos en esos asuntos, presento brevemente algunas preguntas y respuestas que fueron instalándose en el centro de la disciplina a medida que la materialidad de los documentos «multimediales» (el término debe ir encomillado porque uno de sus orígenes es comercial) –que existen mucho antes de los documentos procesados por software– desafiaron el escritorio de trabajo de la historia.

2. ARCHIVOS MULTIMEDIALES, OBJETOS DIGITALES Y EL OFICIO DE LA HISTORIA

La labor con repositorios sonoros y fotográficos puso en evidencia los límites de la concepción tradicional de archivo, centrada como estaba en documentos en papel y registros de Estado. La concepción de «hecho histórico» fue conmovida a medida que nuevas ideas sobre el estatus y las formas de comprensión del registro oral y las fotografías fueron introduciéndose en el trabajo con fuentes. A las ya fuertemente asentadas consideraciones semióticas sobre ese tipo de materiales, se añadieron las exploraciones de la arqueología de medios, las que orientaron su investigación hacia el diagrama, hacia las temporalidades propias de los circuitos e intercambios eléctricos, mucho más allá de las interfaces y la semiosis social (la Música, la Fotografía) (ERNST, 2013). Jussi Parikka muestra cómo, en la actualidad, la crítica de la estela kittleriana en los estudios sobre medios dialoga cada vez mejor con perspectivas ambientales y ecológicas, de modo que puede pensarse una (futura) geología de los medios, en los que los estudios sobre el software y la electrónica compartan laboratorios con la química y la ecología (PARIKKA, 2021). El experimento ocupa un lugar fundamental en esa cantera de análisis de las tecnologías: tanto si llega desde el arte, como forma de anticipación o cortocircuito, cuanto si se obtiene a través del «cacharreo», del experimento con tecnologías en desuso. Un ejemplo será necesario en este punto para observar de qué modo repensamos las mediaciones culturales luego de comprender la agencia de las tecnologías de la comunicación. En uno de sus experimentos, Wolfgang ERNST (2006) reconstruye la «reactivación virtual» del archivo fonográfico de Hornbostel. Para poder ser escuchado, el archivo de grabaciones realizadas hace más de un siglo con gramófonos y cilindros de cera, ha tenido que ser «remediado» a través de una operación óptica (un endoscopio grafica las trazas del cilindro y luego se retraducen esos gráficos al registro sonoro). Para Ernst (2006: 113):

The notion of the archive is in transition, moving towards the audio-visual. As long as there have been archives, the phantasm of recording the acoustically real (i.e., the non-writable) has generated rhetorical, symbolic and scriptural forms of memorizing sound in supplementary ways. Despite the emergence of the phonograph, this new type of record was still subject to forms of inventorization and administration developed in the context of paper-based archives. (Multi-) Media archaeology seeks to reconstruct phantasms of memorizing sound in a pre-technical age and point out the discontinuities which arose with the invasion of audiovisual records into traditional archives, libraries and museums in the twentieth century.

Esa reconstrucción técnica que se adentra en la conversación entre máquinas, tal como lo señala Ernst, conmueve la noción de archivo y, siguiendo argumentos presentes en otros textos suyos, también la relación de la forma archivo con la

memoria social (ERNST, 2018). Pensar las condiciones de producción, circulación y uso de las fuentes documentales siempre ha estado en la agenda de la investigación social. La antropología y la historia no han cesado de tener en consideración la deriva del archivo Hornbostel (GARCÍA, 2012) pero los experimentos de la arqueología de medios introducen nuevas preguntas sobre la producción de la narrativa histórica (¿cambian las condiciones de la escucha, especialmente en su temporalidad, una vez que el material atravesó la traducción lumínica?, ¿nos permite el «cacharreo» técnico de la antigua grabación refigurar las categorías utilizadas en los estudios culturales?, ¿debería hacerlo?).

Por otra parte, el trabajo con objetos digitales produce desafíos que demandan atención. Su abundancia, naturaleza y relación con el mundo continúan discutiéndose desde distintas disciplinas. Su abundancia y su significación son tópicos que la historiografía relaciona constantemente (¿Cómo leer un millón de libros, millones de tweets, etc.?, ¿Cuán significativos resultan los *datasets* sobre conversaciones en redes sociales?) Su materialidad impacta menos en ese campo (como lo hace con la archivística o la filosofía) acaso debido a la todavía poca utilización de objetos digitales propios de la web y las redes, como las páginas web, videos de *Youtube* o performances de *Twitch*. La superposición y complejidad de interfaces convierte a esos objetos en objetos que guardan y reproducen parte de la acción socio-técnica que representan y también algoritmos que habilitan dichas inscripción y reproducción. Un ejemplo apropiado son las páginas web: entendidas como una colección de objetos con diferentes protocolos para organizar su código, algunos elementos están solo referidos en sus «textos» pero son reproducidos en el momento de su lectura/navegación y se modifican de acuerdo con contexto de la ejecución (BRÜGGER, 2012; MILLIGAN, 2019).

A las colecciones y repositorios inmensos se agregan, cada vez más, las donaciones de papeles personales a bibliotecas o archivos que implican computadoras personales, materiales en diskettes, cedés, devedés, etc. ¿Qué elementos son significativos en las computadoras que Salman Rushdie donó a una biblioteca estadounidense?, ¿qué objetos/procesos deben ser conservados y cómo deben ser «leídos»? (CARROLL *et al.*, 2011).

La archivística se ocupa de los objetos digitales desde hace décadas. Algunas perspectivas (MORDELL, 2019) arriesgan un cambio de paradigma, un quinto paradigma (archivos-como-datos o archivos *datificados*), en la línea planteada por Terry Cook en un conocido artículo en donde revisa 150 años de la disciplina (COOK, 2013). Líneas de investigación como la archivística computacional (MARCIANO *et al.*, 2018) o la ingeniería archivística (THIBODEU, 2019) se fundan en desarrollos concretos de la disciplina, pero también en proyectos provenientes de la bibliotecología y la museología. En esas constelaciones interdisciplinarias, la pregunta elemental (*qué es un objeto digital*) considera su basamento filosófico, no exento de tribulaciones, pero también su faceta procedimental, oficiosa. La práctica redefine la pregunta ontológica y permite otras maneras productivas de pensar los objetos digitales. Kenneth THIBODEAU (2002) propuso un modelo para *conservar* objetos digitales, un modelo de capas bajo el paradigma de la base de datos: una capa física, que trata sobre todo al objeto como inscripción (es el campo de la arqueología de medios), una capa lógica, que lo piensa como procesable, ejecutable, etc. Una vez que un objeto se aloja en la memoria de una computadora, el tipo de inscripción deja de ser relevante, y pasa a ser importante la gramática de ese objeto, su codificación, sus relaciones con los programas que procesan esos

datos. El tercer nivel es el conceptual, es el que relaciona al objeto con el mundo (Thibodeau usa el ejemplo de un cajero automático). Lo más interesante de este modelo es que para preservar un objeto digital hay que preservar esas capas y las relaciones entre capas. Son las relaciones entre capas lo que devuelve la significación del objeto (el autor habla de «recrear» para el uso).

La pregunta entonces acerca de cómo podemos interpretar la experiencia ajena a través de los objetos digitales implica la pregunta sobre cómo tenemos que interpretar esos objetos digitales.

3. LOGS DE BBS

En los últimos tiempos son más frecuentes las investigaciones en sede histórica sobre los primeros años de la comunicación mediada por computadoras y los comienzos de la web (BRÜGGER *et al.*, 2017). Esos abordajes se agregan a una lista extensa de investigaciones provenientes de las ciencias sociales que indagan el período formativo de la «era digital». En ese marco sobre tecnologías como el correo electrónico, listas de discusión, USENET, entre otras, son conocidas las pesquisas sobre BBS, como las de Kevin DRISCOLL (2014), Mark HARGADON (2011) y especialmente su presentación pública en el clásico libro de H. RHEINGOLD, *La comunidad virtual* (2012). A su vez, estas aproximaciones cada vez más conectan con historias de las tecnologías y los imaginarios sociales que, insistiendo en el plural de los pliegues, dibujan un ecosistema de casi un siglo de duración, en el que la historia de internet con su mítico origen en ARPANET es solo una de muchas. Una de las contribuciones de la perspectiva histórica a esas *historia(s) de internet* (ABBATE, 2017) que me gustaría subrayar en este apartado es la crítica de las fuentes documentales (por ejemplo, PALOQUE-BERGÈS, 2017). A continuación, intento mostrar cómo el estudio de las continuidades entre las prácticas de la cultura escrita y los por entonces «nuevos medios» puede verse desafiado por las particularidades de la «capa lógica» de los objetos digitales.

Los BBS son sistemas de comunicación que conectan computadoras por medio de un módem y una línea de teléfono. Alguien instala un servidor en su casa y publica su número de teléfono y el horario de funcionamiento. Si quiero conectarme a ese BBS, llamo desde mi computadora a ese número a través de un software, y una vez que se establece la conexión puedo bajar archivos, dejar y leer mensajes, incluso jugar. Los BBS tuvieron su época dorada entre fines de los setenta y principio de los noventa del siglo XX (para Argentina el período fue desde fines de los ochenta y mediados de los noventa). Mantengo el tiempo presente, sin embargo, porque existen en la actualidad algunos BBS en funcionamiento. Mi interés en estudiar la sociabilidad de las primeras comunidades en línea radica en una hipótesis más bien tradicional: las subculturas –el uso del término pretende indicar que el análisis de las comunidades definidas por sus *prácticas significantes* había estado en manos de los estudios culturales– como las que se organizaban alrededor de la astronomía, la ciencia ficción, el anarquismo, el ajedrez, entre muchas otras actividades, podrían haber modificado sus relaciones una vez que construyeron comunidades en línea a través de los BBS. Mi pregunta inicial había sido ¿cómo había cambiado la sociabilidad en esas subculturas a partir del uso de las tecnologías implicadas en los BBS? (Luego comencé a prestar atención a comunidades nacientes como la del *Computer Underground*.)

Inicialmente pensé en dos tipos de fuentes de información: entrevistas orales y *logs* de BBS. Un log es un archivo de «texto plano», es el registro de las conversaciones realizado de modo automático por el software que use la persona que administra el BBS (*sysop*). Un BBS puede guardar cientos o miles de esos registros. Buscar y leer *logs* de BBS fue una de las actividades más importantes en los comienzos de la investigación.

```

Fecha: 08-08-97 (05:09)      Numero: 160 of 171 (Refer# NONE)
Para: All
De: ██████████
Subj: Entrance BBS
Read: NO                    Status: PUBLIC MESSAGE (Echo)
Conf: BBS.ARG (20)         Read Type: GENERAL (+)
En pocas palabras ...
    UU  UU
  UUUUUUUUUU P00Y
UUUUUUUUUU UU00B BUUUUUUUUUUU  UU000 UU000000  UU000000
UUUUUUUU  UU000000
P00B BU0Y000B BUUUUUUUUUUU  UU000BP00B BU00  UU00BU0000
P000BU00Y000B BU0
UUUUUUUUUU UU  P000 0000  P00B  UU000 UU0Y00Y BU00Y000  B
UUUUUUUUUU
000 UU 000 UU00Y P0000 UU000  UU00B BU00Y000  UU00 P000 UU
000 UU
UUUUUUUUUU UU UU00B BUUUUU000  BUUUUUUUUU UU  P000
UUUUUUUU UU000000B
BBBBB BB BB BBB BB BBBBBB B BB BBBB BBBBB
  Llama y averigua como somos y que ofrecemos.(Por supuesto GRATIS)
  eNTRaNCe Soporte oficial de CyberBook
UAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAA
3054-01-581-1990 Lunes a jueves de 19 a 07 Viernes. Sabados, domingos y Feriado
3las 24hs del dia.2 Lineas rotativas de acceso simultaneo al bbs.
3
  AAAAA
    5 8 1 1 9 9 0
  Cincho Ocho Uno Uno Nueve Nueve Cero
---
* Origin: Publica tu chivo mandandolo x net a BBSFIX, 4:900/613 (4:900/261)

```

Figura 1. Mensaje en un log de BBS (Canal sobre ufología, BBSFIX).

Como puede verse (Fig. 1), hay una serie de problemas relacionados con la

interfaz y el software de gestión ejemplificados en el modo en que se imprime el arte ASCII. La disparidad entre lo que se ve en pantalla (Fig. 2) y lo que se lee en una impresión de los registros se advierte rápidamente con esos desplazamientos y transformaciones. La codificación textual y gráfica ocupan buena parte del tiempo del «cacharreo» de los *sysops* (administradores) y algunos análisis de la comunicación mediada por computadoras le dedican, acertadamente, bastante atención (DRISCOLL, 2014; PALOQUE-BERGÈS, 2017).

La pantalla a fines de los ochenta y principios de los noventa era ya una interfaz poderosa, «invisible» cuando era eficiente. Según las personas a las que entrevisté –y lo mismo puede escucharse en el documental sobre los BBS de Jason SCOTT (2005)–, ellas mantenían *conversaciones*. De acuerdo con el *sysop* de *Rolling BBS*, un BBS vía conexión telefónica actualmente en funcionamiento, las personas difícilmente volvían a leer la misma pantalla debido a que cada pantalla tardaba aprox. 30 segundos en refrescar. Tampoco imprimían muchos chats por otras cuestiones técnicas o concretamente porque no tenían impresora.

Son notables los desplazamientos de sentido entre el registro textual (texto plano de los *logs* e impresión del investigador), el registro en pantalla, y la percepción nativa sobre una de las actividades centrales de los comienzos de la comunicación en línea. Pensé que los *logs* eran documentos valiosos y podían ser tratados como fuentes tradicionales (por eso los imprimía) debido a que esas comunidades de finales de los años ochenta y principios de los noventa del siglo XX tuvieron un corazón textual. Sus protagonistas eran sólidos exponentes de la cultura escrita, y sus conversaciones y escritos (como los *ezines*) estaban enraizados en el patrón letrado.

```

Telnet bbs.docksud.com.ar
Subj: La tierra hueca
To: all
From: MHS: [redacted] (TEMP)
Date: Sun Mar 22 1998 07:10:00 UTC

De: Daniel Rosini 800:100/101.15
Fecha: 15 Mar 98 23:07:52
Hola A TODOS!!!

Anduve leyendo un libro de Raymond Bernard sobre la teoria de la tierra hueca
y quer a compartir sus detalles con ustedes...
Confieso que el tema no es nuevo para m .. pero nunca llegu a comprenderlo en su totalidad. En este libro encontr las
respuestas que buscaba...

La teoria de que la esfera terrestre sea hueca, no se formul hace poco... hay trabajos que vienen del siglo pasado,
pero nac en a mediados del siglo 20 se di el caso puntual de una expedici n estadounidense al polo y mas hall ..
Por qu el "m s hall "?
Paso a contarles...

Desde que Col n comprob la redondez de la tierra y que se descubrieron las l neas de fuerza electromagn tica que
rodean a nuestro planeta, se acord que los polos magn ticos eran dos puntos terrestres coincidentes con los polos
geogr ficos donde las l neas de fuerza electromagn ticas flu an hacia la tierra.

M s cerca de nuestro tiempo se comprob que los polos magn ticos no coincid an con los geogr ficos y que se encontraban
a 25 grados de los mismos. Pero he aqu ... que no se dijo toda la verdad, los polos magn ticos no solo no coincid an
con los geogr ficos, sino que variaban en su posici n!!!

[unrapped lines 1-21 of 126]
Reading [ Dock Sud OLD afn.ovni-gral (?-Menu) (10 of 10):

```

Figura 2. Mensaje de un BBS actual, conexión vía telnet
(Canal de ufología, archivo histórico de *BBS DockSud*.)

Pero mientras avanzaba en la lectura de *logs* tenía muy presente los aspectos originales de esas interacciones: el poder de hacer realidad conversaciones a distancia con otras personas, la capacidad de obtener información velozmente, que son algunas de las habilitaciones por las cuales los BBS se pensaron como «comunidades». Pero advertí que, si solo atendía a esas características, podía perder de vista el núcleo letrado de estas comunidades, herederas de una imaginación técnica que tenía al papel como centro de su universo.

Nick MONTFORT (2004) denominó «esencialismo de pantalla» al gesto analítico que olvida la propia historia del software, que tuvo 20, 30 años sin pantalla. Con sus reflexiones sobre software con «interfaz de tinta y papel» (*Eliza/Doctor y Adventure*), él trató de hacer complementarios el enfoque que piensa «la vida en las pantallas» y el enfoque que piensa lo mismo, pero como «vida en el pergamino». El de Montfort es un documento político en que las humanidades digitales van a apoyarse para subrayar la materialidad de lo digital y la necesidad de concebir lo virtual también como continuidad con la cultura escrita. Pero queda claro que la lectura textual de los registros es reductiva y que la representación en «tinta y papel» de lo que se supone fueron conversaciones, aplana la enorme riqueza del complejo sociotécnico de capas de tecnologías de época.

Hay entonces «esencialismo de pantalla» y «esencialismo de imprenta». Lo que aparece con fuerza en este ejemplo es la complejidad de las interfaces y con ello la necesidad de prestar atención a los estudios sobre software. Hay muchas relaciones entre capas para pensar (Fig. 3). Múltiples capas de programación, codificaciones, protocolos, máquinas, pantallas, particularidades de la telefonía, habilitaban ciertas prácticas y dejaban atrás algunas otras. De todas ellas, la naturaleza asincrónica de los intercambios introduce una nueva variación (durante mucho tiempo, la conversación consistió en entrar al BBS y dejar un mensaje, luego había que desocupar la línea telefónica para que otra persona pudiera entrar, y así poder leer los mensajes, dejar los propios). La interacción en diferentes momentos multiplica los contextos de escritura y lectura. Nada de eso quedaba en el «pergamino» de las impresiones que yo hacía al principio de mi investigación.



Figura 3. Tres dimensiones de la comunicación en BBS.

Son muchos los investigadores que han pensado las continuidades y las transformaciones de los objetos tecnológicos y sus capas haciendo hincapié en aspectos que heredan y remedian de tecnologías pasadas (MANOVICH, 2005; BURGESS Y BAYM, 2020; TAYLOR, 2018). Uno de los aspectos más importantes para una perspectiva histórica son las temporalidades específicas que surgen en interacciones entre personas y máquinas. Estuvo claro para mí, al poco tiempo de iniciada la investigación, que la temporalidad había ganado el centro de mis avances. Las temporalidades en sistemas culturales deben ser entendidas como un vector político; es una cuestión relacionada con la imaginación sobre las masas. Beatriz Sarlo, primero inspirada en Raymond Williams y Pierre Bourdieu (SARLO, 1994) y luego en Tomás Maldonado (SARLO, 2018), sostiene que existe una estabilidad entre tecnologías, sectores sociales, géneros literarios y formas de lectura. Para Sarlo, la configuración de principios del siglo XX con centro en la literatura de folletín (textos para leer en el tranvía) logró entrenar a los sectores populares en el dominio de los tiempos de la retórica y los géneros de la prensa gráfica, mientras que la pedagogía de las redes sociales no logra ceder las herramientas para el manejo de los tiempos veloces de los géneros hiperbólicos. Pero ¿cómo podríamos pensar estas interacciones si la ley de archivo (lo que se guarda) y los formatos para archivar los objetos digitales no permiten reponer la complejidad de las interacciones? (la misma pregunta guio a Diana TAYLOR (2016) para pensar las relaciones entre los *escraches* de la organización «Hijos por la Identidad y la Justicia contra el Olvido y el Silencio» y el archivo). Incluso si no acordamos con los postulados últimos de la arqueología de medios a la manera de W. Ernst, hemos de reconocer la importancia de la metodología DIY para la investigación social con tecnologías en desuso, en tanto encontramos la manera de guardar las relaciones entre las capas de los objetos tecnológicos.

Como restos de conversaciones entre personas y computadoras, los *logs* expresan rupturas y continuidades con la cultura escrita, pero su interpretación

no puede realizarse solo con procedimientos tradicionales de la interpretación histórica. El trabajo con tecnologías en desuso es una técnica, en este caso, que permite evitar esencialismos en el análisis de interfaces y habilita una mejor comprensión de las temporalidades inscriptas en *logs* de texto plano.

4. UN ARCHIVO ILEGIBLE: GEOCITIES.COM

Geocities.com fue una empresa fundada en 1994 que permitía publicar páginas web. Pensada como una ciudad, tenía barrios y suburbios temáticos que agrupaban con dificultad creciente tópicos y comunidades (Fig. 4). En 1999 fue adquirida por *Yahoo!* y permaneció, en caída, hasta 2009; aunque en Japón se cerró muy recientemente. En el 2009 un equipo de personas pudo hacer una copia bastante amplia de la ciudad y en la actualidad disponemos de una copia parcial de *Geocities.com* para investigación. Ian Milligan ha escrito sobre la historia de esta colección. Su *History in the Age of Abundance?* (MILLIGAN, 2019) es un texto fundamental sobre las particularidades de una investigación histórica a partir de repositorios de datos nacidos digitales.

¿Cómo recuperar la experiencia popular en la internet de entresiglos? Mi trabajo con la colección de *Geocities.com* no es central en mi investigación, pero a partir del encuentro con ese material comencé a interrogarme acerca de una pregunta axial de las humanidades digitales y su esfuerzo por proponer el trabajo con herramientas de minería de textos y componer lo que Franco MORETTI (2015) denominó «lectura distante».

Se trata de un conjunto enorme de archivos en el que solo los de extensión *htm/html* superan los 15 millones. Son cifras en las que se ha insistido para instalar una obviedad: es muy difícil para los humanos aproximarse al contenido *general* de ese archivo. Solo las notas de un equipo con 7 u 8 años de *trabajo en archivo* ocuparían unas 20 o 30 mil páginas. La pregunta básica de ese esfuerzo sobre este repositorio sin curaduría, *¿cuál es su contenido?* no puede ser respondida en buena medida por la abundancia de lo explorado, pero también por las mediaciones sociales, corporativas y los problemas de la copia y la consulta del investigador.

revisamos el archivo en busca de un solo recurso, por muy significativo que este sea, todo lo que encontraremos serán versiones inferiores de ese recurso, *porque eso es lo que en realidad estamos buscando*» (MORETTI, 2015:106). Se trata de un enfoque que no carece del conocimiento de las metodologías pero que conceptualmente ajusta el alcance de estas (por ejemplo, SCHWANDT, 2020).

Otras intervenciones en favor de las herramientas computacionales para el análisis de textos apostaron por experimentos con mayor agencia de las máquinas (RAMSAY, 2011; JOKERS, 2013). Algoritmos como LDA para el modelado de tópicos (MIMNO, 2013) desafían formas de clasificación humanas a la vez que definen un espacio de interlocución en donde los investigadores han sido también creativos (MIMNO, 2012).

En este punto de la investigación, el momento más angustiante pero más extraordinario de cualquier investigación social, la escena de estar frente a los datos, la escena del momento-inicial de la pesquisa, toda la cuestión para pensar *Geocities.com* pareció dividirse entre establecer límites y trabajar con datos contextualizados, o bien poner a operar algún *software* de minería de datos para «ver qué sale».

Los programas que ejecutan técnicas de las humanidades digitales son cada vez más intuitivos. La complejidad de esas técnicas requiere, de todos modos, habilidades sobre estructuras de datos y codificación. Los proyectos de Ian Milligan, por ejemplo, utilizan un software resultado de uno de sus proyectos (*The Archives Unleashed Project*, <https://archivesunleashed.org/>) pero los cimientos de ese desarrollo son esotéricos para la mayoría de las personas iniciadas en computación. Las bibliotecas de lenguajes como R o Python son más fáciles de usar pero también implican muchas delegaciones, procedimientos «cajanegrizados» y trabajo del tipo «prueba y error».

Lo que quiero indicar con este segundo caso es que, frente a colecciones «ilegibles», la encrucijada interpretativa no está dada por los problemas de las máquinas para comprender la experiencia humana, sino por el problema de las personas para comprender los procedimientos computacionales. En los últimos años eso ha impulsado la conversación sobre metodologías digitales y también el «cacharreo», el uso de esas herramientas. Pero no ha sido igual de notable el interés por el análisis de los algoritmos que usan esos programas. Sin ese análisis, estamos cerca de ser usuarios de instrumentos desconocidos para analizar repositorios desconocidos.

En el procesamiento de lenguaje natural, tanto el pre-procesamiento de los textos cuanto el procesamiento propiamente dicho está saturado de algoritmos dispuestos en una jerarquía estructural compleja. Cada uno de ellos, a su vez, posee una historia (un contexto de producción y contextos de uso y resignificación), y se construyeron en base a decisiones no exentas de ambigüedades. Un ejemplo simple creo que puede iluminar lo que indico aquí. Si quisiera construir una «nube de palabras» (o muchas) para «leer» el contenido de la colección con la que trabajo tendría que interactuar con un sitio web o una aplicación. Tal vez sin saberlo o apenas intuyendo, tendría que interactuar con decenas de algoritmos, solapados en instrucciones o comandos. Antes de «pasar» las palabras para ser procesadas, debo pre-procesar los textos. Además de normalizarlos y estructurarlos tendría que definir cuáles términos son significativos (descartar palabras comunes o infrecuentes/demasiado frecuentes). La mayoría de los algoritmos para preprocesar textos tiene como paso importante el uso de *stopwords*, una lista de

visual propuesto. El espacio tiene variaciones (el movimiento, por ejemplo) y lo tienen también el orden de las etiquetas. La complejidad de las nubes de palabras ha hecho que algunos especialistas hablen de 1era y 2da generación de *tag clouds* (KHUSRO *et al.*, 2018). Los problemas han sido discutidos y existen muchos arreglos a la «nube de palabras» clásica. Se propusieron algoritmos de ordenamientos y clasificación de los términos, nuevos algoritmos de visualización (que es un campo enorme, emparentado con la teoría de grafos) y se discutió también la interpretabilidad y utilización de las nubes.

Debemos conocer al menos los ejes de esa conversación, incluso si continuamos anotando palabras en la lista de *stopwords* luego de sucesivos ensayos, y si continuamos probando diferentes aplicaciones para hacer «nubes de palabras». En los últimos tiempos, desde las humanidades digitales, pero especialmente desde las investigaciones sobre algoritmos, este tipo de vigilancia epistemológica se ha reforzado. Con colecciones como *Geocities.com*, sobre las que no podemos tener una visión de conjunto, no es suficiente el uso de herramientas computacionales: tenemos que adentrarnos en la vida de los algoritmos.

5. CONCLUSIÓN

Este artículo se ocupó de dos escenarios: las sociabilidades en la prehistoria de la web y las colecciones enormes de documentos nacidos digitales. El desafío en el primer escenario produjo la comprensión de que las temporalidades resultado especialmente de la «capa lógica» de los objetos digitales son fundamentales para entender las transformaciones en las subculturas que experimentaron tecnologías por entonces novedosas. Los «esencialismos» en la interpretación de textos me inclinaron a pensar que la experimentación con tecnologías en desuso (*modemismo*, modelos de computadoras como *Commodore 64* o PCs muy «viejas») me pondría más cerca del sentido de la experiencia. La familiaridad de los documentos hallados (*logs* de BBS) derivó en el exotismo de la «remediación» al estilo de la arqueología de medios (un trabajo de laboratorio que la arqueología conoce por estudiar otras prácticas).

En cambio, el desafío del segundo escenario se originó en la pregunta por el contenido de la colección de páginas web, más allá de su puesta en contexto y comprensión en tanto colección, en busca de una *visión de conjunto*, un fotograma propio de la intuición del oficio. Esto dio curso a una indagación sobre los algoritmos (y sus conexiones) que integran las técnicas de las humanidades digitales. La distancia (no familiaridad) con el contenido de los objetos digitales (páginas web de *Geocities.com*) requirió, en este desarrollo, la ayuda de una estrategia clásica de la crítica textual: la crítica de la crítica.

Hace tiempo, Stanley FISH (2012) escribió fuertes ataques contra las humanidades digitales y sus (falsas) promesas. Su defensa de las potencias de la interpretación tradicional, personal, intensa y lenta no tuvieron en cuenta un aspecto elemental de esa operación intelectual humanista: que ella desde siempre estuvo relacionada con tecnologías (desde la toma de notas hasta la edición). La propuesta que se ensaya aquí no es clausurar la conversación con las herramientas computacionales sino avanzar desde dos perspectivas hacia la comprensión de sus *affordances* y de los complejos socio-técnicos donde se construyen: una preocupada por comprender las tecnologías y la otra por criticar las técnicas. Ambos esfuerzos

acompañan a las humanidades desde sus orígenes.

6. AGRADECIMIENTOS

Versiones previas a este artículo han sido discutidas en diferentes eventos académicos. Agradezco especialmente los comentarios y sugerencias de Lila Caimari y Eden Medina. Asimismo, agradezco las indicaciones de quienes evaluaron anónimamente este texto para su versión definitiva.

7. REFERENCIAS

- ABBATE, J. (2017): «What and where is the Internet? (Re)defining Internet histories», en *Internet Histories*, 1: 8-14.
- BAILEY, J. (2013): «Disrespect des Fonds: Rethinking Arrangement and Description in Born-Digital Archives», *Archive Journal*, 3. URL: <https://www.archivejournal.net/essays/disrespect-des-fonds-rethinking-arrangement-and-description-in-born-digital-archives/>.
- BRUGGER, N. (2012): «Web history and the web as a historical source», *Zeithistorische Forschungen*, 9 (2): 316-325.
- BRÜGGER, N.; GOGGIN, G.; MILLIGAN, I.; SCHAUER, V. (2017): «Introduction: Internet histories», en *Internet Histories*, 1: 1-7.
- BURGESS, J.; BAYM, N. (2020): *Twitter: A Biography*, NYU Press, New York.
- CAIMARI, L. (2017): *La vida en el archivo: goces, tedios y desvíos en el oficio de la historia, Siglo Veintiuno*, Buenos Aires.
- CARROLL, L.; FARR, E.; HORNSBY, P.; RANKER, B. (2011): «A Comprehensive Approach to Born-Digital Archives», *Archivaria* 72: 61-92.
- COOK, T. (2013): «Evidence, memory, identity, and community: four shifting archival paradigms», *Archival Science*, 13: 95-120.
- DRISCOLL, K. (2014): *Hobbyist inter-networking and the Popular Internet Imaginary: forgotten histories of networked personal computing, 1978-1998*, Tesis de Doctorado (Comunicaciones), University of Southern California.
- CRYMBLE, A. (2021): *Technology and the Historian. Transformations in the Digital Age*, University of Illinois Press, Urbana.
- ERNST, W. (2006): «Dis/continuities. Does the Archive Become Metaphorical in Multi-Media Space? », en W. CHUN y T. KEENAN (eds.), *New Media, Old Media: A History and Theory Reader*, Routledge, New York: 105-123.
- ERNST, W. (2013): *Digital Memory and the Archive*, UMP, Minneapolis-London.
- ERNST, W. (2018): «El archivo como metáfora. Del espacio de archivo al tiempo de archivo», *Nimio* 5 (11):31-42.
- FISH, S. (2012): «Mind Your P's and B's: The Digital Humanities and Interpretation», *New York Times*, 23 de enero. <https://opinionator.blogs.nytimes.com/2012/01/23/mind-your-ps-and-bs-the-digital-humanities-and-interpretation/>
- GARCÍA, M. Á. (2012): «Las músicas de Tierra del Fuego en su versión (etno) musicológica», en K. BODEMER (coord.), *Cultura, sociedad y democracia en América Latina, Iberoamericana-Vervuert*, Madrid: 285-301.
- GINZBURG, C. (2004): «Conversar con Orion», en *Tentativas, Prohistoria*, Rosario:

321-336.

- HARGADON, M. (2011): *Like City Lights, Receding: ANSi Artwork and the Digital Underground, 1985-2000*, Tesis de maestría, Concordia University.
- JOCKERS, M. (2013): *Macroanalysis: Digital Methods and Literary History*, University of Illinois Press, Urbana.
- KHUSRO, Sh. et al. (2018): «Tag Clouds: Past, Present and Future», en *Proceedings of the National Academy of Sciences, India Section A: Physical Sciences*: 369-381.
- LUHN, H. P. (1958): «The Automatic Creation of Literature Abstracts», *IBM Journal of Research and Development*, 2 (2): 159-165.
- MANOVICH, L. (2005): *El lenguaje de los nuevos medios de comunicación: La imagen en la era digital*, Paidós, Barcelona.
- MARCIANO, R.; LEMIEUX, V.; HEDGES, M.; ESTEVA, M.; UNDERWOOD, W.; KURTZ, M.; CONRAD, M. (2018): «Archival Records and Training in the Age of Big Data», en J. PERCELL et al., *Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education*, Emerald Publishing, Bingley: 179-199.
- MILLIGAN, I. (2019): *History in the Age of Abundance?: How the Web Is Transforming Historical Research*, McGill-Queen's University Press, Ottawa.
- MIMNO, D. (2012): «Reconstructing pompeian households», *arXiv preprint*, <https://arxiv.org/abs/1202.3747>.
- MIMNO, D. (2013): «A Wrapper around the Java Machine Learning Tool MALLET Reference Manual», en CRAN. <https://cran.r-project.org/web/packages/mallet/mallet.pdf>
- MONTFORT, N. (2004): «Continuous Paper. The Early Materiality and Workings of Electronic Literature». https://nickm.com/writing/essays/continuous_paper_mla.html
- MORDELL, D. (2019): «Critical Questions for Archives As (Big) Data», *Archivaria*, 87:140-161.
- MORETTI, F. (2015): *Lectura distante*, FCE, Buenos Aires.
- PALOQUE-BERGÈS, C. (2017): «Usenet as a web archive. Multi-layered archives of computer mediated», en N. BRÜGGER. (ed.), *Web 25: histories from the first 25 years*, Peter Lang, New York: 229-252.
- PARIKKA, J. (2021): *Una geología de los medios*, Caja Negra, Buenos Aires.
- PETERS, B. (2016): *Digital keywords: A vocabulary of information society and culture*, Princeton University Press, Princeton.
- PUTNAM, L. (2016): «The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast», *The American Historical Review*, 121 (2): 376-402.
- RAMSAY, S. (2011): *Reading Machines: Toward an Algorithmic Criticism*, University of Illinois Press, Urbana.
- RHEINGOLD, H. (2012): *La comunidad virtual. Una Sociedad sin fronteras*, Gedisa, Madrid.
- SARLO, B. (1994): *El imperio de los sentimientos. Narraciones de circulación periódica en la Argentina (1917-1927)*, Siglo Veintiuno, Buenos Aires.
- SARLO, B. (2018): *La intimidad pública*, Seix Barral, Buenos Aires.
- SCHWANDT, S. (2020): «Métodos digitales para la semántica histórica», *Conceptos históricos*, 8 (5):160-196.
- SCOTT, J. (2005): *BBS: The Documentary* [DVD] URL: <https://archive.org/details/BBS.The.Documentary>
- TAYLOR, D. (2016): *El archivo y el repertorio. El cuerpo y la memoria cultural en las Américas*, Ediciones Universidad Alberto Hurtado, Santiago de Chile.

- TAYLOR, T. L. (2018): *Watch Me Play: Twitch and the Rise of Game Live Streaming*, Princeton University Press, Princeton.
- THIBODEAU, K. (2002): «Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years», en *The State of Digital Preservation: An International Perspective*, Conference Proceedings, IIS, URL: <https://www.clir.org/pubs/reports/pub107/thibodeau/>
- THIBODEAU, K. (2019), «The Construction of the Past: Towards a Theory for Knowing the Past», *Information*, 10 (11), 332. <https://doi.org/10.3390/info10110332>.
- WINTERS, J. (2020), «Web Archives and (digital) history: a troubled past and a promising future? », en N. BRÜGGER e I. MILLIGAN, *The SAGE Handbook of Web History*, SAGE, New York-London: 593-606.

