

# Lectura distante, fragmentada y colaborativa en el archivo infinito

*A Distant, Fragmented and Collaborative Reading in an Infinite Archive*

**Jairo Antonio Melo Flórez**

EL COLEGIO DE MICHOACÁN, jairom@colmich.edu.mx

Este ensayo se centra en las posibilidades que brindan las aplicaciones informáticas para la lectura de textos en un contexto de datos masivos; es decir, de un archivo infinito, lo cual es un aspecto clave de las humanidades digitales. El problema por revisar es hasta qué punto una aplicación informática puede leer textos y traducirlos para la comprensión humana. Para enfrentar este propósito se considera pertinente abordar tres formas de lectura que están relacionadas intrínsecamente con la revolución informática: distante, fragmentada y colaborativa. El análisis de estas formas de lectura implica cuestiones más amplias sobre la lectura en la web y la dicotomía entre las posibilidades que pueden brindar las aplicaciones informáticas y los resultados empíricos de la utilización de modelos computarizados en el estudio de la cultura.

**PALABRAS CLAVE:** lectura, información, humanidades digitales, datos, análisis de textos.

This essay focuses on the possibilities that the informatics applications provide for texts reading in a context of big data, or in other words in an infinite archive, which is a key issue for digital humanities. The central problem is to what extent an informatics application is able to read texts and to make them comprehensive for humans. In order to address this problem, it is necessary to consider three forms of reading directly relate with the digital turn: distant, fragmented, and collaborative. The analysis of these forms of reading involves broader questions about reading in the web and the dichotomy between the possibilities that the informatics applications and the empiric results of using computerized models in the study of culture.

**KEYWORDS:** Reading, information, digital humanities, data, text-analysis.

La traducción y reciente publicación en Argentina del libro *Lectura distante* de Franco Moretti,<sup>1</sup> publicado en inglés en el año 2013, constata el interés que viene generando el área de las humanidades digitales en América Latina. Moretti se ha convertido en una de las figuras visibles dentro del campo de los estudios histórico-literarios, en especial, después de publicar su afamado y en cierta medida polémico *Graphs, Maps and Trees: Abstract Models for Literary History*, en el cual propone el análisis cuantitativo de los textos literarios en periodos de larga duración. Moretti pretende trascender propuestas de lectura cercana (*Close Reading*) como el análisis de contenido<sup>2</sup> y procura llevar los estudios literarios hacia los grandes volúmenes de texto al aplicar una lectura no literal sino cuantitativa de grandes volúmenes de información que permitan superar la interpretación y en cambio hacer evidentes estructuras, entendidas como repeticiones, regularidades, órdenes y patrones que se hacen visibles al percibir la literatura desde la *longue durée*. En sus palabras: “the models I have presented [graphs, maps and trees] also share a clear preference for explanation over interpretation; or perhaps, better, for the explanation of general structures over the interpretation of individual texts”.<sup>3</sup> Apreciaciones como las de Moretti constatan el sentido cuantitativo y no narrativo de ciertas tendencias en las humanidades digitales y ratifican el recelo por un posible enfoque neopositivista<sup>4</sup> o por lo menos de una nue-

<sup>1</sup> Franco Moretti, *Lectura distante*, trad. Lilia Mosconi (Buenos Aires: Fondo de Cultura Económica, 2015).

<sup>2</sup> Klaus Krippendorff, *Content analysis: an introduction to its methodology*, 2ª ed. (Thousand Oaks: Sage, 2004).

<sup>3</sup> Franco Moretti, *Graphs, Maps, Trees: Abstract Models for Literary History* (Londres, Nueva York: Verso, 2007), 91. Es interesante cómo Moretti recupera el sentido dado por Dilthey al acercamiento que se puede adoptar en relación con un texto, esto es como explicación, en el sentido de una aplicación del modelo de las ciencias naturales a las ciencias históricas (enfoque positivista) o como interpretación, entendida como la actitud fundamental de las ciencias humanas ante los textos. Paul Ricœur, *From Text to Action*, *Essays in Hermeneutics 2* (Evanston: Northwestern University Press, 1991), 105.

<sup>4</sup> Son muchos los artículos, ensayos y mensajes de blog que han tratado el sentido neopositivista de las humanidades digitales entre los cuales destaco el texto de Johanna Drucker, “Humanistic Theory and Digital Scholarship”, en *Debates in the Digital Humanities*, ed. Matthew Gold, 85-95 (Minneapolis: Universidad de Minnesota, 2012). <http://dhdebates.gc.cuny.edu/debates/text/34>. También es de resaltar el comentario he-

va versión de la cliometría, esta vez en forma de *culturomics*, un análisis cuantitativo de la cultura que busca fenómenos culturales perceptibles después de “minar” grandes volúmenes de libros digitalizados.<sup>5</sup>

A pesar de los esfuerzos de Moretti y su equipo del *Stanford Literary Lab*,<sup>6</sup> los resultados no han sido como esperaba, por lo menos así lo manifestó en una entrevista al diario argentino *La Nación* en la cual dijo: “Lo que tratamos de responder es si al estudiar grandes archivos de textos con estos nuevos métodos ha cambiado nuestra visión de la historia. Honestamente debo decir que, hasta ahora, eso no ha sucedido. Por eso, este método es por ahora más una promesa que una realidad”.<sup>7</sup> Esta declaración de Moretti evidencia la dicotomía entre las capacidades de las aplicaciones informáticas y las posibilidades explicativas que se pueden derivar de su uso. No hay duda que con un algoritmo se podría “leer” una cantidad de textos más allá de la capacidad de cualquier ser humano, pero a pesar de su capacidad para extraer y relacionar contenidos aún no se ha encontrado la manera en que el programa entienda la forma del texto, por ejemplo, si las oraciones tienen un sentido literal o irónico.<sup>8</sup>

Después de una década de experimentación, el avance en el uso de herramientas informáticas para la lectura de datos masivos entre los historiadores es significativo, tal como se evidencia en la reciente publicación de Shawn Graham, Ian Milligan y Scott Weingart, *Exploring Big Historical Data. The Historian's Macroscope*.<sup>9</sup>

---

cho por Anacleto Pons al enfoque de Moretti en Anacleto Pons, *El desorden digital: guía para historiadores y humanistas* (Madrid: Siglo XXI, 2013), 125-26.

<sup>5</sup> Jean-Baptiste Michel *et al.*, “Quantitative Analysis of Culture Using Millions of Digitized Books”, *Science* (16 de diciembre de 2010): 176-182. doi:10.1126/science.1199644.

<sup>6</sup> “Stanford Literary Lab-Directed by Franco Moretti and Mark Algee-Hewitt”, <http://litlab.stanford.edu/> (Fecha de consulta: 29 de mayo de 2016).

<sup>7</sup> Natalia Blanc, “Franco Moretti: ‘El estudio de la cultura perdió mucho por no seguir un método científico’”, *La Nación*, jueves 17 de septiembre de 2015. <http://www.lanacion.com.ar/1828487-franco-moretti-el-estudio-de-la-cultura-perdio-mucho-por-no-seguir-un-metodo-cientifico> (Fecha de consulta: 30 de mayo de 2016).

<sup>8</sup> Esta crítica a la “minería de datos” proviene de “Franco Moretti”.

<sup>9</sup> Shawn Graham, Ian Milligan y Scott Weingart, *Exploring Big Historical Data. The Historian's Macroscope* (Londres: Imperial College Press, 2016).

Hay que entender asimismo que para un historiador, como en general para cualquier humanista, tratar con datos masivos no significa lo mismo que para un analista de mercado u otro ingeniero que trabaja con *big data*. Siguiendo la explicación dada por Shawn Graham, los datos masivos pueden comprenderse como una cantidad de información más grande de lo que normalmente un investigador podría abarcar en su trabajo cotidiano;<sup>10</sup> es decir, que es humanamente imposible leer e interpretar de manera tradicional.

Para el científico de las humanidades, el contenido de la información suele encontrarse de forma semiestructurada, es decir, catalogada de tal manera que es posible identificar el tipo de documento, las características físicas del original, el lugar de procedencia, autores, e incluso un resumen o descriptor del contenido; todo ello es lo que se denomina como metadatos, es decir, información que permite ubicar, ordenar y guardar datos de manera automatizada. Pero hay mucha información que no está estructurada y consiste en una inmensa colección de unidades de lenguaje para cuyo análisis no es posible fiarse de la automatización, por ello en muchas ocasiones se prefiere construir las bases de datos antes que “minar” los recursos Web para recolectar información útil para resolver los problemas planteados por las ciencias humanas. Así, las humanidades digitales se han enfrentado a lo que puede considerarse una nueva necesidad de lectura en el sentido de un novel paradigma de interpretación de textos históricos denominada la lectura distante.

## LECTURA DISTANTE

Una lectura distante es de hecho una no lectura, así definió Martin Mueller a la acción de permitir que las máquinas lean textos.<sup>11</sup> Sin

<sup>10</sup> Shawn Graham, *Exploring Big Historical Data-The Historian's Macroscope Book Launch @Carleton\_U* (Ottawa, 2015), <https://www.youtube.com/watch?v=NjrZAJeD-VYA>. Desde una perspectiva de la ingeniería de la información el *big data* representa un conjunto de datos cuyo tamaño supera la capacidad de captura, almacenaje, gestión y análisis de un software estándar de análisis estadístico. Chris Snijders, Uwe Matzat y Ulf-Dietrich Reips, “‘Big Data’: Big Gaps of Knowledge in the Field of Internet Science”, *International Journal of Internet Science* 7(1) (2012): 1.

<sup>11</sup> Martin Mueller, “Notes towards a user manual of Monk-MONK-GSLIS Wiki”,

embargo, a pesar de la desafortunada concepción de la lectura automatizada como *not-reading* (años después Mueller modificó el término por *Scalable Reading*),<sup>12</sup> su propuesta señala la consecución de un anhelo que ha estado presente desde hace mucho tiempo: la posibilidad de leer todos los libros del mundo. En términos prácticos implica quizás una aspiración más prosaica que consiste en superar la inevitable atadura a una disciplina acumulativa en la que los aportes más significativos al conocimiento son realizados por los más ancianos, tan arcaica es ante los ojos posmodernos nuestra ciencia.

El tiempo de lectura es inversamente proporcional a las exigencias productivas de la actual academia y en este factor puede radicar el interés que suscitan las formas de lectura digital, sin embargo, la propuesta de lectura cuantitativa con la aplicación de técnicas como la minería de datos y el análisis de contenido imponen retos a la misma manera en que leemos los textos. Un ejemplo práctico se puede observar al usar la herramienta *Google Books Ngram Viewer*, en la cual se puede “leer” la recurrencia de una frase en la totalidad de libros disponibles en Google Books por año y a través de un periodo de tiempo específico. En este caso, la no lectura de millones de libros en español brinda el siguiente resultado que se muestra en la figura 1.

La figura 1 es una consulta simple que muestra las relaciones entre tres palabras que representan cada una un concepto: justicia, perdón y castigo. El ejercicio de leer el diagrama puede ser sencillo y podría reflejar un significativo interés por la “justicia” hacia la segunda mitad del siglo xvii y durante prácticamente los dos siglos que van de 1750 a 1950; la tarea por seguir podría ser explorar los lotes de libros de esos años o realizar operaciones combinadas entre n-gramas. Por ejemplo, si en lugar de buscar por cada concepto por separado realizamos la pesquisa de la relación entre justicia+castigo y justicia+perdón el resultado cambia significativamente como se observa en la figura 2.

<http://web.archive.org/web/20100713210851/https://apps.lis.illinois.edu/wiki/display/MONK/Notes+to+wards+a+user+manual+of+Monk> (Fecha de consulta: 21 de junio de 2016).

<sup>12</sup> Martin Mueller, “Scalable Reading”, *Scalable Reading* on May 29, 2012, [https://scalablereading.northwestern.edu/?page\\_id=22](https://scalablereading.northwestern.edu/?page_id=22).

FIGURA 1. N-grama por palabras en el lote de libros de *Google Books*

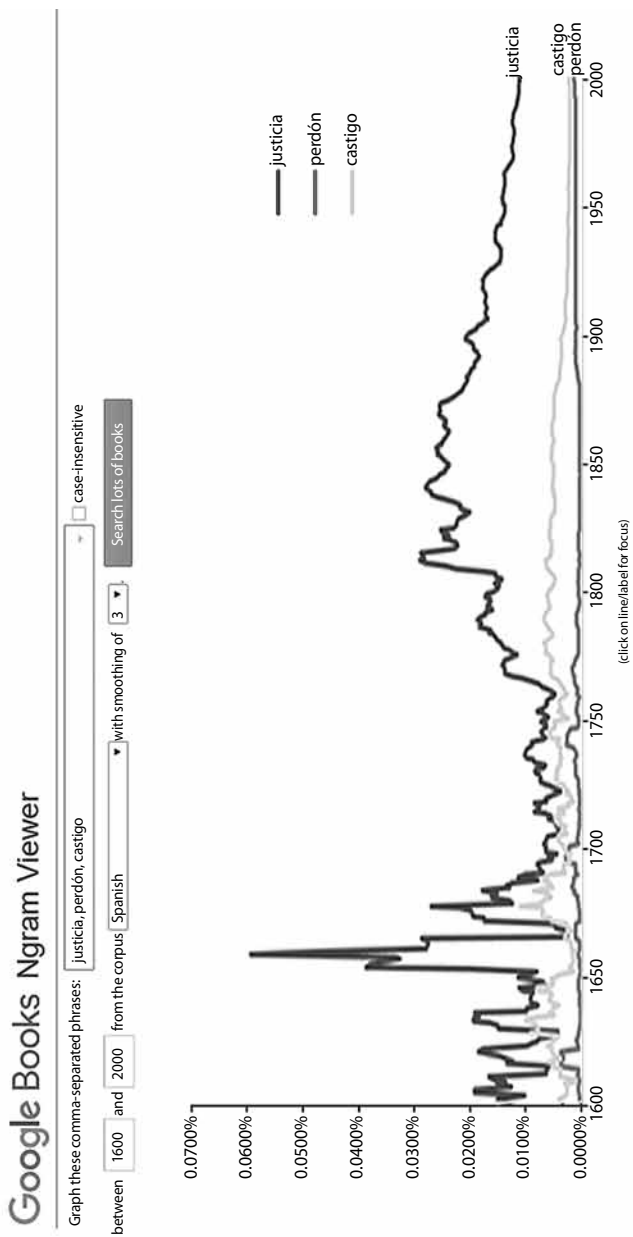
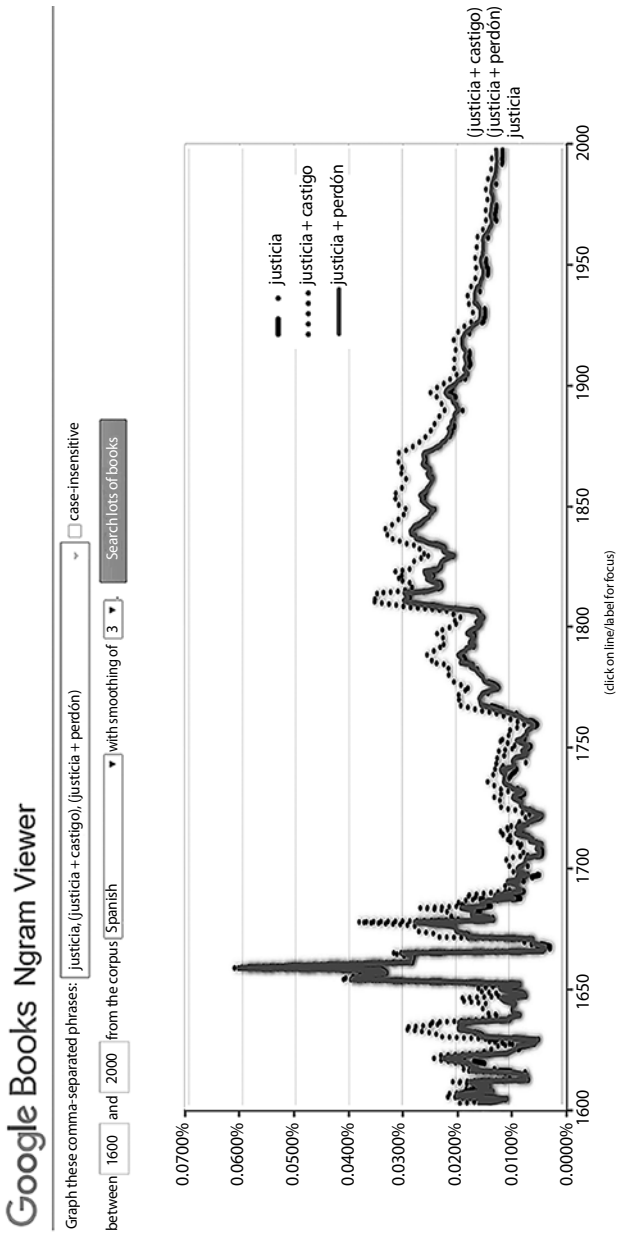


FIGURA 2. N-grama por palabras relacionadas en el lote de libros de *Google Books*



La similitud de ambas tendencias se debe a que la palabra dominante (justicia) se suma con la variable lingüística que se añade, en este caso perdón o castigo. Aunque son los mismos resultados que se obtuvieron para la figura 1, la lectura de esta gráfica hace más evidentes la importancia que se le brinda al castigo con relación a la justicia desde el siglo xvii hasta la primera mitad del siglo xx. Sin embargo esta lectura sigue siendo artificiosa ya que de alguna manera se lee lo que las variables quieren que se grafique (en otras palabras, el lector dirige los resultados de la lectura). Pero una muestra de lectura realizada por la máquina se puede observar con la aplicación de un “comodín”, esto es, un carácter especial que le indica a la máquina que debe realizar una lectura específica del corpus textual cuyo resultado es desconocido para el usuario lector. En la figura 3 se observa el uso de dicha técnica.

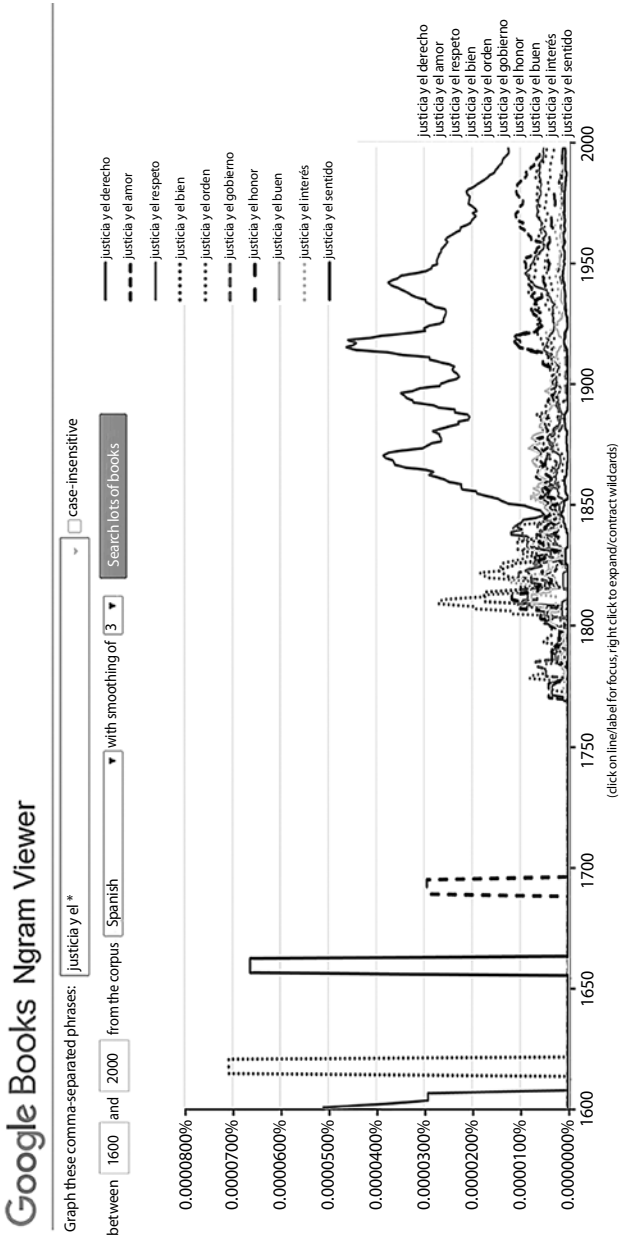
En este caso, figura 3, ya no es el lector quien indica a la máquina exactamente qué buscar sino es la máquina la que lee por el usuario e indica cuáles son las unidades de lenguaje más recurrentes en el lote de libros durante el tiempo consultado. Por ejemplo, me indica la presencia de la idea de “la justicia y el bien” en un lote de libros que va de 1613 a 1620, de “la justicia y el sentido” entre 1656 y 1661, y una interesante tendencia que inicia en 1768 y que terminará dominando hacia mediados del siglo xix como es la frase “la justicia y el derecho”. Claramente esta “lectura” genera muchos interrogantes y campos de observación, como el periodo de 1770 a 1850, lo cual evidencia de alguna manera la eclosión de ideas que se vivió en el mundo hispanohablante en la época.

El ejercicio puede seguir durante horas, sin embargo, la posibilidad de llegar a explicaciones por vía de la máquina son limitadas. En primer lugar, no es posible la generación de *n-grams* con una longitud mayor a cinco unidades del lenguaje, es decir, difícilmente se alcanza a formar una frase y con ello interpretaciones más sesudas de las tendencias e ideas en el tiempo. Para ello se deberá descargar el *Raw data* puesto a disposición por Google y con ese insumo hacer análisis de gran escala.<sup>13</sup>

<sup>13</sup> “Datasets”, *Google Ngram Viewer*, <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html> (Fecha de consulta: 20 de junio de 2016).



FIGURA 3. N-grama generado con el uso de un “comodín”



Otra opción es recurrir al *topic modeling*, esto es, definir “temas” con los cuales puedan hallarse sentidos “ocultos” en un gran contenido de texto. Un ejemplo de esta técnica se puede observar en el proyecto *Mining the Dispatch* (realizado por Robert K. Nelson, director del *Digital Scholarship Lab* en la universidad de Richmond) en el cual se observó la tendencia de ciertos temas en los textos publicados por el periódico *Richmond Daily Dispatch* que se evidenciaron gracias al uso del programa MALLET (*MAchine Learning for Language Toolkit*).<sup>14</sup> Gracias al uso de esta aplicación se encontró una serie de palabras que a su vez se pudo relacionar con contenidos específicos tales como anuncios, diatribas u opiniones, con los cuales fue posible trabajar para luego insertar los temas en un contexto social y narrativo. En todo caso Nelson es cauto respecto al uso del *topic modeling* y, en general, de la lectura distante, y dice al respecto: “topic modeling is certainly not a replacement for conventional, close reading methods. It and other distant reading methods do, however, provides historians an additional method that allows us to examine and detect patterns within not a sampling but in the entirety of an archive”.<sup>15</sup>

Es evidente que la lectura distante brinda una manera particular de “leer” la información digital que es imposible de abarcar por el humano tradicional, sin embargo, es importante mencionar que por el momento forma parte de la caja de herramientas de un porcentaje muy pequeño de especialistas. Entre más especializadas son las herramientas menos humanistas tienen la formación, el tiempo y la paciencia necesaria para entender un lenguaje tan ajeno al narrativo. A pesar de ello es claro que el uso de bases de datos y búsquedas es una acción cotidiana, tan sólo que dicha práctica se ha interiorizado de tal manera que parece natural que se hagan búsquedas en servicios para la comunidad científica y a partir de ello se realice un tipo específico de lectura fragmentada.

<sup>14</sup> Robert K. Nelson, “Mining the Dispatch”, <http://dsl.richmond.edu/dispatch/pages/intro> (Fecha de consulta: 20 de junio de 2016); “MALLET homepage”, <http://mallet.cs.umass.edu/> (Fecha de consulta: 20 de junio de 2016).

<sup>15</sup> Nelson, “Mining the Dispatch”.

## LECTURA FRAGMENTADA

Como el mismo Moretti lo ha manifestado, la lectura distante es más un anhelo que una realidad. En un reciente artículo Lara Putnam<sup>16</sup> llama la atención de que a pesar de las grandes herramientas analíticas que están disponibles para leer la información digital vía gráficas, mapas o árboles, la gran mayoría de humanistas (específicamente los historiadores) aún realizamos las pesquisas de información vía Google, Google Books, JSTOR, FamilySearch, entre otros lugares, en búsqueda de datos cualitativos, esto es por temas, personas, momentos o lugares. El *digital turn* en este sentido se convierte en una práctica cotidiana de los historiadores y humanistas de manera casi imperceptible, algo que ya había resaltado Anacleto Pons cuando dijo

todos nos hemos digitalizado de manera informal, de modo que escribimos con procesadores de texto, nos comunicamos por correo electrónico, consultamos información en los buscadores, etcétera. Sin embargo, tratamos ese mundo como si sólo fuese “un apéndice, una curiosidad, una distracción, algo superfluo”, que poco o nada tiene que ver con nuestro “verdadero trabajo”.<sup>17</sup>

La crítica de Putnam consiste en señalar que el interés de las humanidades digitales, en especial, aquellos interesados en el trabajo con datos masivos, se ha centrado en el campo de lo posible (la lectura distante) en lugar de transformar lo más sencillo e informal: la búsqueda de información.<sup>18</sup>

Si la lectura distante y el modelado de datos plantean tantos problemas para el historiador y otros investigadores del campo de las humanidades ¿por qué aferrarse a lo posible antes que a lo ya alcanzado? Volviendo a Putnam, las búsquedas de información digital se

<sup>16</sup> Lara Putnam, “The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast”, *American Historical Review* 121(2) (2016): 377-402

<sup>17</sup> Anacleto Pons, “‘Guardar como’. La historia y las fuentes digitales”, *Historia Crítica* (43) (2011): 41.

<sup>18</sup> Putnam, “The Transnational”, 378-79.

han vuelto tan intuitivas que contrastan con la complejidad y cantidad de información disponible en repositorios masivos de información. Sin embargo, lo verdaderamente revolucionario no estaría marcado por el complejo modelado de datos desestructurados sino en el acercamiento al conocimiento transnacional gracias a la lectura desestructurada de información que, no sólo está descuadrada, sino que perdió su arraigo a un espacio nacional y se ancla ahora en el ciberespacio.<sup>19</sup>

Tal vez no exista duda acerca del cambio de escala de análisis temporal y espacial que configuró el aumento de recursos disponibles al investigador en la Web. Considero que la historia digital es ante todo una historia global<sup>20</sup> y como tal debería ser la lectura realizada por el investigador. Es decir, las búsquedas por categorías cualitativas que se realizan diariamente expresan relaciones que son halladas de manera automática por el algoritmo de búsqueda. Además entre más complejo sea el algoritmo y mejor almacenada esté la información en el repositorio más evidentes resultarán dichas relaciones. De lo anterior surge un reto para los repositorios nacionales que consiste en lograr que los motores de búsqueda más avanzados puedan indexar su información o en su lugar mejorar las herramientas disponibles de búsqueda, lo cual es más complejo y sin duda requiere el trabajo de un equipo de desarrolladores especializados.

De manera cotidiana los investigadores estamos leyendo en red a través de buscadores y, por ende, lo hacemos de manera fragmentada. Cualquier estudiante está familiarizado con la lectura dispersa para la resolución de tareas en las cuales el buscador jerarquiza las fuentes según la pertinencia y cantidad de información disponible de acuerdo con una serie de palabras claves o preguntas recurrentes. Google Books hace la tarea incluso más sencilla al haber indexado el contenido completo de sus libros, lo cual permite señalar el lugar exacto en el que las categorías de búsqueda se encuentran ubicadas en cada libro, ordenando los resultados de acuerdo con la recurrencia de las claves marcadas en la pesquisa. De esta manera el buscador

<sup>19</sup> *Ibid.*, p. 380.

<sup>20</sup> Jairo Antonio Melo Flórez, "Historia digital: la memoria en el archivo infinito", *Historia Crítica* (43) (2011): 83-84, doi:10.7440/histcrit43.2011.06.

“lee” y selecciona por nosotros, quienes tenemos la tarea posterior de revisar la pertinencia de lo hallado a partir del contexto de las obras, ese pequeño paso que muchos estudiantes dejan de lado por el afán del copiar y pegar.

La lectura fragmentada no es una novedad, un estudio de 1999 realizado en la Universidad de Antioquia (Colombia) demostró que los estudiantes leían de manera literal y fragmentada debido al uso continuado de fotocopias,<sup>21</sup> lo que se esperaría de alguna manera se viese superado por la accesibilidad a textos completos no se vio correspondido por una reacción docente a la lectura digital. Pero ésta no es una consecuencia de las fotocopias ni de la era digital, la lectura fragmentada, según Roger Chartier, fue favorecida por la popularización del códex y el desarrollo de nuevos gestos como el establecer índices, hojear el libro, citar pasajes, entre otros.<sup>22</sup> Sin embargo, el texto descuadrado del mundo digital promueve una lectura aún más dispersa: “Esta lectura discontinua y segmentada que supone y produce, según la expresión de Umberto Eco, una ‘alfabetizzazione distratta’, es una lectura rápida, fragmentada, que busca informaciones y no se detiene en la comprensión de las obras, en su coherencia y totalidad”.<sup>23</sup>

Es bastante conocido el pasaje de *Así habló Zaratustra* donde el ermitaño señala que la sabiduría se consigue al rumiar (*Wiederkäuhen*) los textos; para Nietzsche, el ritmo acelerado de la edad moderna impedía el ejercicio de pensar. En efecto, rumiar es cada día más complicado, si bien, las búsquedas permiten el acceso a miles de fragmentos que contienen una categoría (o concepto o frase) que los conecta. Dicha interconexión no se hace de manera consciente, aún más en sistemas como Google que por motivos empresariales se ven en la necesidad de ocultar sus algoritmos al usuario. Si las humanidades digitales contribuyen en este aspecto es precisamente en la “revelación” del código, sobre todo, porque la gran ma-

<sup>21</sup> Luz Stella Castañeda N. *et al.*, “La lectura en la Universidad de Antioquia: Informe preliminar”, *Revista Signos* 32(45-46) (1999): 83-101, doi:10.4067/S0718-09341999000100010.

<sup>22</sup> Roger Chartier, “¿La muerte del libro?”, *Co-herencia* 4(7) (2007): 126.

<sup>23</sup> *Ibid.*, p. 127.

yoría de los entusiastas de las *digital humanities* son promotores del código abierto, del acceso abierto, del software libre y del aprendizaje de lenguajes de programación.

La profundización en métodos de búsqueda relacional, no sólo en *big data*, sino en repositorios de cualquier dimensión, es una tarea que puede brindar efectos benéficos para la hermenéutica de textos históricos, sin embargo, no permitirá escapar, por lo menos en el corto plazo, a las tentaciones del anacronismo que una categoría aplicada en una línea de tiempo pueda generar. La interpretación es una actividad humana que hasta ahora no ha podido ser sistematizada, por lo cual se podría esperar que otras formas de lectura masiva con mayor intervención de actores conscientes podrían combatir la fragmentación textual de la era digital.

#### LECTURA COLABORATIVA

Otra de las innovaciones provenientes del campo de las humanidades digitales es el denominado *crowdsourcing*, otro término importado de la ingeniería de mercados en el cual varias personas vinculadas o externas a una empresa se comprometían voluntariamente a buscar la solución a un problema. En las humanidades el término ha mutado para comprender la vinculación de voluntarios en la ciencia, por ejemplo, subiendo documentos a la red, participando experiencias, e incluso colaborando con tareas como la transcripción de manuscritos.<sup>24</sup> Los ejemplos más representativos de *crowdsourcing* se encuentran en proyectos como Wikipedia, el proyecto Gutenberg o el Internet Archive, los cuales surgen de iniciativas privadas sin ánimo de lucro, con una mayor capacidad de maniobra que una institución gubernamental o que un gigante industrial como Google.

Algunos investigadores han visto en el *crowdsourcing* una posibilidad para construir conocimiento. Una de estas iniciativas fue el

<sup>24</sup> Stuart Dunn, "More than a business model: crowd-sourcing and impact in the humanities", *Impact of Social Sciences*, 21 de marzo de 2013, <http://blogs.lse.ac.uk/impactofsocialsciences/2013/03/21/more-than-a-business-model-crowd-sourcing-and-impact-in-the-humanities/>

proyecto *What Do You Do With A Million Readers?*, el cual surgió en respuesta a la pregunta problemática de Gregory Crane *What Do You Do with a Million Books?* La respuesta original de Crane consistió en que en caso de tener un millón de libros digitales lo más apropiado sería convertir los textos mediante un OCR, procesarlos al idioma deseado mediante un traductor automatizado y finalmente extraer la información mediante técnicas de minería de datos.<sup>25</sup> La propuesta del grupo de la UCLA conformado por Roja Bandari, Timothy Tangherlini y Vwani Roychowdhury desarrolló una perspectiva diferente. Estos investigadores encontraron que las humanidades digitales se habían enfocado en explorar los recursos digitalizados por repositorios como Google Books mediante técnicas como la lectura distante, pero habían pasado por alto la explosión de lectores que comentan los libros en foros en línea y otros sitios de acceso abierto. En lugar de “minar” los libros lo hicieron con las reseñas voluntarias del sitio Goodreads para lo cual escogieron los cinco libros más populares (con más de 500,000 evaluaciones) de los cuales obtuvieron un estimado de 3,000 reseñas. El análisis de la información recolectada se hizo mediante la búsqueda de relaciones entre nombres, lugares, objetos y conceptos claves de cada novela (para lo cual se valieron de las guías de estudio *SparkNotes*). Finalmente fue posible encontrar algunas maneras en las que los lectores se relacionaban con la lectura sin que se pudiera llegar a conclusiones definitivas al respecto.<sup>26</sup>

Si bien, los resultados son nuevamente preliminares brindan otra estrategia de acercamiento a la lectura de grandes volúmenes de información, claro está, cuando es posible incentivar la retroalimentación por parte del público. Tal vez el *feedback* sea un problema manifiesto en nuestro contexto ya que en caso de tener un millón de lectores no necesariamente significa una misma cantidad de comen-

<sup>25</sup> Gregory Crane, “What Do You Do with a Million Books?”, *D-Lib Magazine* 12(3) (marzo de 2006), doi:10.1045/march2006-crane.

<sup>26</sup> Roja Bandari, Timothy Roland Tangherlini y Vwani Roychowdhury, “What Do You Do With A Million Readers?”, [http://web.archive.org/save/\\_embed/http://dh2015.org/abstracts/xml/TANGHERLINI\\_Timothy\\_Roland\\_What\\_Do\\_You\\_Do\\_With\\_A\\_/TANGHERLINI\\_Timothy\\_Roland\\_What\\_Do\\_You\\_Do\\_With\\_A\\_Millio.html](http://web.archive.org/save/_embed/http://dh2015.org/abstracts/xml/TANGHERLINI_Timothy_Roland_What_Do_You_Do_With_A_/TANGHERLINI_Timothy_Roland_What_Do_You_Do_With_A_Millio.html)

tarios. Para traer a colación un proyecto particular, en el año 2012 construimos junto a Miguel Darío Cuadros un sitio titulado “Archivo de publicidad colombiana 1800-1950”,<sup>27</sup> el cual surge ante la cierta popularidad que contaban algunos grupos en Facebook que compartían anuncios publicitarios que los administradores hallaban en los periódicos y revistas que guardaban en sus hogares. La idea consistía en conformar un repositorio con anuncios que difícilmente podían tener los usuarios en sus casas y por ello decidimos convocar a la comunidad de estudiantes, historiadores y aficionados a la historia para que compartieran aquellos avisos publicitarios que encontraran en medio de sus pesquisas hemerográficas. A través del tiempo pudimos acumular entre algunos colaboradores una pequeña cantidad de anuncios (211 en total), de los cuales requeríamos dos tareas básicas: transcribir y describir. De nuevo recurrimos a la comunidad a través de diferentes canales pero, con excepción de la contribución de la historiadora María Fernanda Erazo Obando y del equipo de trabajo del proyecto, no hubo otra contribución significativa.

El caso del Archivo Histórico de Publicidad tal vez no sea representativo de la generalidad de proyectos de historia digital en América Latina, pero sí permite evidenciar la dificultad que existe para lograr la contribución voluntaria del público y la comunidad académica, ambos capturados por las redes sociales. La evaluación que se hace en este aspecto es que a pesar del humilde avance de las humanidades digitales la relación con las aplicaciones digitales se encuentra aún en un nivel de consumo. El profesor de periodismo Dennis Jerz brinda una definición interesante de las humanidades digitales que ataca directamente este asunto: “We apply technology because we must participate in digital culture in order to understand it. Full participation in digital culture means contributing to (creating) the cultural economy, not simply observing (consuming)”.<sup>28</sup>

<sup>27</sup> “Archivo de Publicidad Colombiana 1800-1950”, <http://apc.historiaabierta.org/> (Fecha de consulta: 20 de mayo de 2016).

<sup>28</sup> “How do you define Humanities Computing / Digital Humanities? - Taporwiki”, [http://www.artsrn.ualberta.ca/taporwiki/index.php/How\\_do\\_you\\_define\\_Humanities\\_Computing/\\_Digital\\_Humanities%3F](http://www.artsrn.ualberta.ca/taporwiki/index.php/How_do_you_define_Humanities_Computing/_Digital_Humanities%3F) (Fecha de consulta: 17 de junio de 2016).



Es evidente asimismo que los repositorios institucionales no están contribuyendo a la generación de una cultura de lectura colaborativa ya que casi ninguno contiene un módulo de trabajo colectivo, ya sea del tipo *Scripto* que permite la transcripción voluntaria de documentos<sup>29</sup> o tan sencillos como casillas de comentarios, foros de discusión o espacios para reseñas. El *feedback*, por lo general, se encuentra encasillado en las redes sociales (Facebook y Twitter de preferencia) o en los comentarios de soporte que no están disponibles al público; esto hace que exista una relación unidireccional, muy en el sentido del “servicio al cliente”, pero no en la generación de un *crowdreading*. La construcción de propuestas de lectura colectiva de fuentes, de interpretación de documentos y el análisis de esa participación “masiva” es definitivamente un campo que puede ampliar significativamente la perspectiva de la historia digital y de las humanidades digitales en general,<sup>30</sup> aprovechando un comportamiento desarrollado en la Web 2.0 (la necesidad de participar, comentar, reaccionar y compartir) y así superar el brindar repositorios anclados en la Web estática.<sup>31</sup>

#### LECTURAS, NO LECTURAS, INTERPRETACIONES, NARRATIVAS

Si algo queda en evidencia en este ensayo es que la historia y las humanidades digitales han brindado una nueva problematización a la hermenéutica histórica, algo que de cierta manera se había dejado como ejercicio de los filósofos y que, gracias a los giros lingüístico y digital, se ha devuelto al campo de discusión del humanista. Es complejo, sin embargo, desembarazarse de la sospecha de positivismo que subyace en las formas de lectura de las humanidades

<sup>29</sup> “Scripto”, consultado el 17 de junio de 2016, <http://scripto.org/>

<sup>30</sup> Aunque hay que señalar que en los estudios literarios la ventaja es significativa, como queda en evidencia con la propuesta de Javier del Puerto en <http://crowdreading.com/es>

<sup>31</sup> Al respecto es interesante el ensayo de Sheila A. Brennan y T. Mills Kelly, “Why Collecting History Online is Web 1.5”, Roy Rosenzweig Center for History and New Media, *Essays on History and New Media* (2009), <http://chnm.gmu.edu/essays-on-history-new-media/essays/?essayid=47>

digitales, más cuando autores como Franco Moretti prácticamente eliminan la posibilidad interpretativa de los textos en pos de unas ciencias humanas más cercanas a las ciencias duras. Sin embargo, esto no es más que un espejismo y los débiles resultados de la experimentación cuantitativa así lo demuestran hasta ahora. Lo que es realmente relevante son las nuevas formas de lectura y, por lo tanto, de interpretación que brinda el mundo digital.

Scott B. Weingart señala que la posibilidad de que la historia digital brinde nuevas explicaciones a viejas preguntas o genere nuevas preguntas está limitada en gran medida a las condiciones que se presentan desde el plano institucional. En ciertas ocasiones lo que lograría la historia digital es llegar a los mismos resultados con un método diferente, algo que señala Weingart genera cierta reticencia a los métodos computacionales por no brindar “nuevas conclusiones” (aunque paradójicamente ratifica viejas respuestas).<sup>32</sup> Sin embargo, ¿es la historia digital un medio para la narrativa histórica o es en sí misma una nueva narrativa? Es difícil saberlo aún, la mayoría de acercamientos a este campo se han hecho desde el plano metodológico o como una adaptación de la lógica de la ingeniería a la investigación histórica, pero todavía no queda claro cómo puede contribuir a los elementos no narrativos de formación de sentido histórico y con ello contribuir a una crisis paradigmática que pueda llevar a nuevas preguntas y, en consecuencia, a reinterpretaciones de relato histórico.<sup>33</sup>

De cierta manera los historiadores y humanistas digitales nos encontramos como los inquisidores de la biblioteca de Babel, proponiendo métodos para desvelar los secretos de ese hexágono infinito que aunque no es el universo sí contiene una parte muy importante de lo que fuimos y somos como especie. Como el eterno viajero que atraviesa la biblioteca en cualquier dirección, hallamos al final que las posibilidades de encontrar patrones son inconmensurables y a la vez

<sup>32</sup> Scott Weingart, “‘Digital History’ Can Never Be New”, *The Scottbot Irregular*, 2 de mayo de 2016, <http://scottbot.net/digital-history-can-never-be-new/>

<sup>33</sup> Acerca del paradigma y la pregunta en el sentido de la historia: Jörn Rüsen, *Tiempo en ruptura*, Colección humanidades, trad. Christian Sperling (México: Universidad Autónoma Metropolitana, 2014), 173-180.

finitas, y en cada viaje algorítmico por los corredores, escaleras y hexágonos de la red nos topamos con “el mismo desorden (que, repetido, sería un orden: el Orden)”.<sup>34</sup>

## BIBLIOGRAFÍA

- ARCHIVO DE PUBLICIDAD COLOMBIANA 1800-1950. <http://apc.historiaabierta.org/> (Fecha de consulta: 20 de mayo de 2016).
- BANDARI, Roja, Timothy Roland TANGHERLINI y Vwani ROYCHOWDHURY. “What Do You Do With A Million Readers?” [http://web.archive.org/save/\\_embed/http://dh2015.org/abstracts/xml/TANGHERLINI\\_Timothy\\_Roland\\_What\\_Do\\_You\\_Do\\_With\\_A\\_/TANGHERLINI\\_Timothy\\_Roland\\_What\\_Do\\_You\\_Do\\_With\\_A\\_Millio.html](http://web.archive.org/save/_embed/http://dh2015.org/abstracts/xml/TANGHERLINI_Timothy_Roland_What_Do_You_Do_With_A_/TANGHERLINI_Timothy_Roland_What_Do_You_Do_With_A_Millio.html)
- BLANC, Natalia. “Franco Moretti: ‘El estudio de la cultura perdió mucho por no seguir un método científico’”. *La Nación*, jueves 17 de septiembre de 2015 <http://www.lanacion.com.ar/1828487-franco-moretti-el-estudio-de-la-cultura-perdio-mucho-por-no-seguir-un-metodo-cientifico> (Fecha de consulta: 30 de mayo de 2016).
- BORGES, Jorge Luis. “La biblioteca de Babel”. En *Ficciones, El Aleph, El informe de Brodie*, 36-41. Caracas: Fundación Biblioteca Ayacucho, 1986.
- BRENNAN, Sheila A. y T. MILLS KELLY. “Why Collecting History Online is Web 1.5”. Roy Rosenzweig Center for History and New Media. *Essays on History and New Media*, 2009. <http://chnm.gmu.edu/essays-on-history-new-media/essays/?essayid=47>
- CASTAÑEDA N., Luz Stella y José Ignacio HENAO S. “La lectura en la Universidad de Antioquía: Informe preliminar”. *Revista Signos* 32(45-46) (1999): 83-101 doi:10.4067/S0718-09341999000100010.
- CHARTIER, Roger. “¿La muerte del libro?”. *Co-herencia* 4(7) (2007): 119-29.

<sup>34</sup> Jorge Luis Borges, “La biblioteca de Babel”, en *Ficciones, El Aleph, El informe de Brodie* (Caracas: Fundación Biblioteca Ayacucho, 1986), 41.

- CRANE, Gregory. "What Do You Do with a Million Books?". *D-Lib Magazine* 12(3) (marzo de 2006) doi:10.1045/march2006-crane. "Datasets". *Google Ngram Viewer*. <http://storage.googleapis.com/books/ngrams/books/datasetv2.html> (Fecha de consulta: 20 de junio de 2016).
- DRUCKER, Johanna. "Humanistic Theory and Digital Scholarship". En *Debates in the Digital Humanities*, ed. Matthew Gold, 85-95. Minneapolis: Universidad de Minnesota, 2012. <http://dhdebates.gc.cuny.edu/debates/text/34>
- DUNN, Stuart. "More than a business model: crowd-sourcing and impact in the humanities". *Impact of Social Sciences*, 21 de marzo de 2013. <http://blogs.lse.ac.uk/impactofsocialsciences/2013/03/21/more-than-a-business-model-crowd-sourcing-and-impact-in-the-humanities/>
- GRAHAM, Shawn. *Exploring Big Historical Data-The Historian's Macrocope Book Launch @Carleton\_U*. Ottawa, 2015. <https://www.youtube.com/watch?v=NjrZAJeDVYA>
- GRAHAM, Shawn, Ian MILLIGAN y Scott WEINGART. *Exploring Big Historical Data. The Historian's Macrocope*. Londres: Imperial College Press, 2016.
- "How do you define Humanities Computing / Digital Humanities? - Taporwiki". [http://www.artsrn.ualberta.ca/taporwiki/index.php/How\\_do\\_you\\_define\\_Humanities\\_Computing/\\_Digital\\_Humanities%3F](http://www.artsrn.ualberta.ca/taporwiki/index.php/How_do_you_define_Humanities_Computing/_Digital_Humanities%3F) (Fecha de consulta: 17 de junio de 2016).
- KRIPPENDORFF, Klaus. *Content analysis: an introduction to its methodology*. 2ª ed. Thousand Oaks: Sage, 2004.
- "MALLET homepage". <http://mallet.cs.umass.edu/> (Fecha de consulta: 20 de junio de 2016).
- MELO FLÓREZ, Jairo Antonio. "Historia digital: la memoria en el archivo infinito". *Historia Crítica* (43) (2011): 82-103 doi:10.7440/histcrit43.2011.06.
- MICHEL, Jean-Baptiste, Yuan KUI SHEN, Aviva P. AIDEN, Adrian VERRIS, Matthew K. GRAY, The Google Books Team, Joseph P. PICKETT *et al.* "Quantitative Analysis of Culture Using Millions of Digitized Books". *Science* (16 de diciembre de 2010): 176-182. doi:10.1126/science.1199644.

- MORETTI, Franco. *Graphs, Maps, Trees: Abstract Models for Literary History*. Londres y Nueva York: Verso, 2007.
- \_\_\_\_\_. *Lectura distante*. Trad. Lilia Mosconi. Buenos Aires: Fondo de Cultura Económica, 2015.
- MUELLER, Martin. “Notes towards a user manual of Monk-MONK-GSLIS Wiki”. <http://web.archive.org/web/20100713210851/https://apps.lis.illinois.edu/wiki/display/MONK/Notes+towards+a+user+manual+of+Monk> (Fecha de consulta: 21 de junio de 2016).
- \_\_\_\_\_. “Scalable Reading”. *Scalable Reading*, 29 de mayo de 2012. [https://scalablereading.northwestern.edu/?page\\_id=22](https://scalablereading.northwestern.edu/?page_id=22)
- NELSON, Robert K. “Mining the Dispatch”. <http://dsl.richmond.edu/dispatch/pages/intro> (Fecha de consulta: 20 de junio de 2016).
- PONS, Anacleto. *El desorden digital: guía para historiadores y humanistas*. Madrid: Siglo XXI, 2013.
- \_\_\_\_\_. “‘Guardar como’. La historia y las fuentes digitales”. *Historia Crítica* (43) (2011): 38-61.
- PUTNAM, Lara. “The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast”. *American Historical Review* 121(2) (2016): 377-402.
- RICŒUR, Paul. *From Text to Action*. Essays in Hermeneutics 2. Evanston: Northwestern University Press, 1991.
- RÜSEN, Jörn. *Tiempo en ruptura*. Colección humanidades. Trad. Christian Sperling. México: Universidad Autónoma Metropolitana, 2014.
- “Scripto”. <http://scripto.org/> (Fecha de consulta: 17 de junio de 2016).
- SNIJDERS, Chris, Uwe MATZAT y Ulf-Dietrich REIPS. “‘Big Data’: Big Gaps of Knowledge in the Field of Internet Science”. *International Journal of Internet Science* 7(1) (2012): 1-5.
- “Stanford Literary Lab-Directed by Franco Moretti and Mark Algee-Hewitt”. <http://litlab.stanford.edu/> (Fecha de consulta: 29 de mayo de 2016).
- WEINGART, Scott. “‘Digital History’ Can Never Be New”. *The Scottbot Irregular*, 2 de mayo de 2016. <http://scottbot.net/digital-history-can-never-be-new/>