RESEÑA *EXPLORING BIG HISTORICAL DATA. THE HISTORIAN'S MACROSCOPE*

Víctor Gayol
Colegio de Michoacán, A.C.
vgayol@colmich.edu.mx

Una de las facetas más interesantes y productivas de la historia y de las Humanidades Digitales (HD) es la posibilidad de manejar y analizar grandes cantidades de información procesada mediante recursos informáticos. De hecho, el inicio de las HD —llamadas hace muchos años con el nombre más humilde de *informática aplicada a las humanidades*— se puede ubicar en el año de 1946 con el arranque del proyecto monumental del padre Roberto Busa SJ, el *Index Thomisticus*, en el que se propuso recopilar y analizar la frecuencia, concordancia y contexto de cada palabra del léxico contenido en la obra de Santo Tomás y autores relacionados. Conocido desde hace décadas como el inventor de la concordancia generada por computadora, Busa reunió y analizó un total de 13 millones de registros entre 1946 y 1974; proyecto que se puso en línea en 2005[1]. Aquellos inicios fueron difíciles pues requirieron de máquinas muy caras y sofisticadas que solamente había en los grandes centros de investigación o en las empresas corporativas, así como una ingente cantidad de horas/hombre de programación, procesamiento y análisis.

Setenta años después, desde sus propias computadoras personales y echando mano de software libre de código abierto, los alumnos de un doctorado en historia pueden repetir la experiencia de Busa e, incluso ir más allá, pues pueden procesar analíticamente un corpus documental con 127 millones de palabras como las que contiene la colección de 197 mil juicios criminales que fueron recopilados del *Old Bailey* londinense[2]. Esto es, abordar una cantidad ingente de textos documentales (*big data*) y procesarlos extrayendo el total del conjunto de datos para determinar las constantes y las variables presentes en los discursos, utilizar para ello herramientas de visualización gráfica, organizar los patrones en forma de modelos tópicos (temáticos) del total del corpus documental o utilizar análisis de redes para encontrar, por ejemplo, las conexiones semánticas entretejidas en esos modelos discursivos. Para quien quiera ahondar en la cultura jurídica y la relación entre los discursos de la criminalidad y el pecado en el siglo XVIII tardío, por ejemplo, la aplicación de estos recursos le puede ser de gran utilidad.

Exploring Big Historical Data aborda estos y otros muy interesantes ejemplos de aplicación de recursos informáticos en la investigación histórica y humanística, pues es precisamente un manual pensado para estudiantes de posgrado en esas disciplinas que necesiten analizar el discurso encerrado en corpus documentales a gran escala. En siete capítulos, los autores demuestran cómo aplicar los recursos informáticos que pueden estar a la mano de cualquier estudiante —es decir, sin grandes recursos económicos—, para el procesamiento de *big data*. Y lo didáctico comienza desde el principio al definir qué significa una inmensidad de datos para un historiador o un humanista. Los analistas de datos computacionales quizá se mofarían de nuestras definiciones; pero para un alumno de estudios literarios cien

1 http://www.corpusthomisticum.org/it/index.age

2 http://www.oldbaileyonline.org

novelas del tamaño de *Guerra y Paz* presupone una inmensidad textual inabarcable. Como dicen los propios autores refiriéndose a un viejo proverbio —la belleza está en el ojo del espectador—, cada disciplina definirá qué es inmenso para su forma de hacer las cosas.

Exploring Big Historical Data contiene una serie de aproximaciones al uso creativo de recursos que están en la web, en la que se alojan repositorios inmensos como el de más de 80 terabytes de *Internet Archive*, el uso de herramientas de gestión y análisis (como Zotero) y otros más que se pueden crear desde nuestras propias computadoras con un poco de aprendizaje de lenguajes de programación de alto nivel, como Python, para crear breves *plugins* que interactúen con la información digitalizada en línea. Lo que es más interesante en este libro es la manera creativa con la que se va proponiendo el proceso de integración de todas estas herramientas para un único fin: mejorar los procesos de investigación con herramientas y recursos digitales. Y, sobre todo, el manual es una invitación a convertirse en un *historiador o humanista programador*. Quizá esto es lo que podría horrorizar a muchos colegas y estudiantes de posgrado, pero programar en lenguajes como Python no resulta tan difícil a la larga, y la posibilidad que ofrece aprender programación para integrar varias herramientas para el análisis en nuestra investigación muy bien merece el esfuerzo.

Otra característica interesante del libro es la forma en la que fue construido. Se trata de una obra colaborativa compuesta entre Shawn Graham (Carleton University), Ian Milligan (University of Waterloo) y Scott Weingart (Indiana University), que recurrió desde el principio a la discusión e intercambio de los textos en plataformas de Internet. El borrador de *Exploring Big Historical Data* comenzó a escribirse en línea y abierto a la lectura pública de cualquiera que se interesase en el proyecto y quisiera enviar retroalimentación a los autores. Por supuesto que el proyecto terminado es sensiblemente diferente del borrador que aún está en línea y consultable, *The Historian's Macroscope: Big Digital History*,[3] y sirve como un complemento de la publicación formal. En este sitio web se puede seguir mucho del proceso de escritura y corrección del manual, lo cual lo hace un documento muy interesante.

Además, el libro y el sitio web están directamente interrelacionados con otros dos proyectos colaborativos de aprendizaje de programación; utilización de recursos informáticos y presentación de resultados de investigación para historiadores y humanistas en línea: *The Programming Historian*[4] y el libro de Daniel Cohen y Roy Rosenzweig, *Digital History: A Guide to Gathering*, *Preserving, and Presenting the Past on the Web*.[5] El primero es un sitio creado por un equipo de trabajo de investigadores y estudiantes que ofrece manuales para

3 http://www.themacroscope.org/2.0/

4 http://programminghistorian.org

5 Daniel J. Cohen y Roy Rosenzweig (2005), Digital History: A Guide to Gathering, Preserving, and presenting the Past on the Web, Philadelphia, University of Pensylvania Press: http://chnm.gmu.edu/digitalhistory/

la creación y utilización de recursos informáticos para historiadores desde 2012. El segundo es un libro ya clásico sobre el potencial que tiene Internet para la difusión del conocimiento histórico. El uso de estos dos en conjunto con el manual de Graham, Milligan y Weingart potencia provechosamente el aprendizaje. Un manual excelente y entretenido de técnicas y recursos que no estaría mal comenzar a utilizar en México y en toda América Latina.

REFERENCIA

Graham, S., Milligan, I. y Weingart, S. (2016). *Exploring Big Historical Data. The Historian's Macroscope*. London: Imperial College Press.