



Amoxtli
ISSN: 0719-997X
amoxtli@uft.cl
Universidad Finis Terrae
Chile

El uso de la minería de datos como heurística para la teoría de la historia y la historia de la historiografía

Rota, Alesson Ramon

El uso de la minería de datos como heurística para la teoría de la historia y la historia de la historiografía

Amoxtli, núm. 7, 2021

Universidad Finis Terrae, Chile

Disponible en: <https://www.redalyc.org/articulo.oa?id=615769609002>

DOI: <https://doi.org/10.38123/amox7.205>

El uso de la minería de datos como heurística para la teoría de la historia y la historia de la historiografía

The use of data mining as a heuristic for the theory of history and history of historiography

Alesson Ramon Rota *

Universidade de Campinas, Brasil

alesson.rota@gmail.com

 <https://orcid.org/0000-0001-9167-7903>

 <http://buscatextual.cnpq.br/buscatextual/visualizacv.do?id=3896360099405972>

DOI: <https://doi.org/10.38123/amox7.205>

Redalyc: <https://www.redalyc.org/articulo.oa?id=615769609002>

Recepción: 01 Enero 2022

Aprobación: 30 Marzo 2022

RESUMEN:

El objetivo de este artículo se centra en presentar Crono.Data, que es una interfaz para visualizar conceptos, constelaciones de autores y citas indexadas de colecciones digitales. Crono.Data se está desarrollando a partir de la investigación "Red intelectual y panamericanismo: relaciones entre la política y la escritura de la historia en el Cono Sur (1922-1940)", que tiene como objetivo estudiar una constelación de autores americanistas (panamericanos y latinoamericanos) que organizaron congresos de historia, publicaron libros sobre Historia de América y articularon la fundación de instituciones y su financiamiento, además de debatir cuestiones teóricas y metodológicas sobre la escritura de la historia y la educación. Actualmente, indexamos las colecciones del Instituto Histórico Geográfico (1838-2008), la Biblioteca Brasileira Guita y José Mindlin (incompleta), la Biblioteca de Obras Raras de la UFRJ, la Biblioteca Nacional (solo *Revista Americana* de Itamaraty) y la Academia Nacional de Historia de Argentina (1924 a 1948). Utilizamos métodos de minería de datos para encontrar la información que consideramos relevante para la investigación, y técnicas de modelado de datos para transformarla en nuevas formas de representación y visualización. Se presentaron formas de construir la base de datos y formas de analizar la información encontrada.

PALABRAS CLAVE: historiografía, minería de datos, representación, humanidades digitales.

ABSTRACT:

The object of this article is focused on presenting Crono.Data, which is a visual interface to visualize concepts, author constellations and indexed citations from digital collections. Crono.Data is being developed from the research "Intellectual Network and Pan-Americanism: Relations between politics and the writing of history in the Southern Cone (1922-1940)", which aims to study a constellation of Americanist authors (Pan-American and Latin American), who organized history congresses, published books on the subject of American History, and articulated the founding of institutions and financing, as well as debating theoretical and methodological issues on the writing of history and education. Currently, we have indexed the collections of the Instituto Histórico Geográfico (1838-2008), the Biblioteca Brasileira Guita and José Mindlin (incomplete), the Library of Rare Works of the UFRJ, the National Library (only *Revista Americana* of Itamaraty) and the Academia Nacional de History of Argentina (1924 to 1948). We used data mining methods to find the information that we consider relevant for research, and data modeling techniques to transform it into new forms of representation and visualization. Ways to build the database and ways to analyze the are presented.

KEYWORDS: historiography, data mining, representation, digital humanities.

NOTAS DE AUTOR

* Doctorando en Historia con financiamiento FAPESP.

INTRODUCCIÓN 1

El uso de tecnologías para el análisis de datos e información ha sido uno de los campos de la informática que más han crecido en los últimos años, según encuestas y proyecciones de Oracle, multinacional especializada en bases de datos.² En 2002 apareció el *Data Science Journal*, publicado por el International Council for Science. En los primeros volúmenes se destaca la publicación de artículos centrados en el tratamiento de la información para la medicina, la biología, la industria y las ciencias políticas.³ A pesar de que la ciencia de datos surge como una especialización de las áreas de estadística y minería de datos, siendo una subárea de la informática, su crecimiento en los últimos años está asociado al aumento constante de datos producidos en formato digital y a la necesidad de poder analizarlos. También se observa que varias áreas de la investigación han hecho uso de dicho conocimiento, además de las matemáticas, la informática o la ingeniería, así como las ciencias sociales y el periodismo.

Como escribieron De Castro y Ferrari, la minería de datos consiste en producir conocimiento a partir de información extraída de bases de datos.⁴ Según el trabajo de Mehmed Kantardzic, la minería de datos se basa en el conocimiento clásico de las matemáticas y las estadísticas actualizadas mediante el uso de la informática, a través de la inteligencia artificial y el aprendizaje automático.⁵ Aunque las definiciones parecen ser técnicas, demostraré cómo emplear el análisis textual basado en el estudio de algunas herramientas. Quizás la analogía del concepto de *minería* ayude más en esta fase inicial, porque se trata de la búsqueda de objetos valiosos cuyo hallazgo da un arduo trabajo.

Las intersecciones entre la informática y las humanidades se explorarán más a fondo a través de lo que se ha discutido como humanidades digitales. La propuesta es derivar de este debate argumentos que puedan ser utilizados para pensar en la escritura de la historia sustentada en métodos computacionales. Se asume que el trabajo del historiador se basa en la hermenéutica y que, por tanto, los métodos computacionales no deben reemplazar los métodos ya conocidos en historiografía, sino interpretar otras formas de análisis o extracción de información. A modo de ejemplo, presentaremos la herramienta denominada Crono.Data, que se está desarrollando a partir de la investigación "Red intelectual y panamericanismo: relaciones entre la política y la escritura de la historia en el Cono Sur (1922-1940)".⁶

Crono.Data es una herramienta de visualización de conceptos, constelaciones de autores y citas construidas a partir de colecciones digitales en línea, como el Instituto Histórico Geográfico (1838-2008), la Biblioteca Brasiliana Guita y José Mindlin (incompleta), la Biblioteca de Obras Raras de la UFRJ, la Biblioteca Nacional (solo *Revista Americana* de Itamaraty) y la Academia Nacional de Historia de Argentina (1924 a 1948). La premisa es sumar el mayor número de colecciones disponibles, pero como aún no existe un equipo de trabajo colectivo, los índices se realizan según la relación con el objeto de investigación doctoral. Se entiende que Crono.Data puede ser una herramienta útil para otras investigaciones, ya que mapea conceptos clásicos y autores debatidos en las áreas de teoría de la historia, historia de la historiografía e historia intelectual, o puede ser pensado como un ejemplo de cómo construir una base de datos para usar en búsquedas históricas.

HUMANIDADES DIGITALES Y HERMENÉUTICA

Durante el siglo XX hubo varias iniciativas que utilizaron la informática como herramienta para la investigación en humanidades. Un caso conocido fueron los estudios de Roberto Busa, quien en alianza con IBM en la década de 1940 creó una plataforma capaz de clasificar palabras, frases y citas en los libros de Tomás de Aquino o en los que se refieren a él a partir del siglo XIII.⁷ En las últimas décadas, la computación ha dado numerosos subsidios a la estilometría, un campo de estudio centrado en la detección de patrones en los lenguajes, que permite el reconocimiento de autoría en textos apócrifos.⁸ En 1991 el

término “giro computacional” ya aparece en la colección de ensayos editada por el filósofo Leslie Burkholder.⁹ La relación entre la informática y las disciplinas humanísticas, como la literatura y las ciencias sociales, ha suscitado innumerables debates. Por un lado, hay autores como Franco Moretti y Matthew Jockers que ven la informática como la posibilidad de una nueva forma de representación de la escritura que abre interconexiones a escala global.¹⁰ Por otro lado, autores como Alexander Galloway y Nan D. Z. ven problemas en reducir las cuestiones literarias al uso de la informática.

La historia del concepto de “humanidades digitales” revela de alguna manera la complejidad de los debates. Desde 1989, la Alianza de Organizaciones de Humanidades Digitales (ADHO) organiza congresos con el fin de discutir las relaciones entre la informática y las humanidades. En ese momento se utilizó el concepto de computación de las humanidades, por lo que la mayoría de los estudios partieron desde la computación hacia los objetos de estudio. Según Allington et al., los datos extraídos computacionalmente se usaron en una analítica fría, que buscaba remover temas contemporáneos y urgentes, como género, raza, clase, sexualidad, entre otros.¹¹ En 2001 ADHO organizó el Seminario Curricular de Humanidades Digitales, que de alguna manera marca un nuevo vocabulario en el debate. Si analizamos el gráfico elaborado a partir de publicaciones en inglés disponibles en Google Books (Figura 1), notamos un cambio en el uso de la terminología.

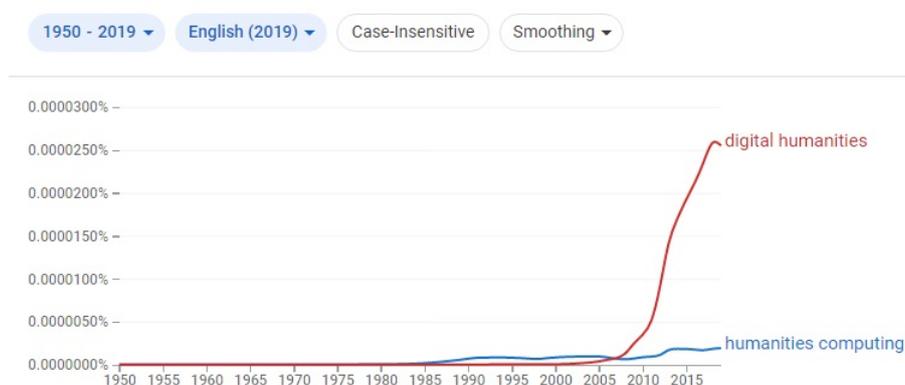


FIGURA 1
Comparación de frecuencia de los conceptos “humanities computing”
y “digital humanities” según la colección de Google Books

Google Ngram. Disponible en <https://books.google.com/ngrams/graph?content=humanities+computing%2Cdigital+humanities>

En el año 2012, The Humanities and Technology Camp publicó un manifiesto que define las humanidades digitales a través de reflexiones sobre la transición de lo analógico a lo digital, haciendo referencia al conjunto de las ciencias humanas y sociales, las artes y letras.¹² Las definiciones en torno a las humanidades digitales son abiertas y asumen la relación transdisciplinar entre diferentes áreas del conocimiento. Esta amplitud se puede resumir en el proyecto dirigido por Geoffrey Rockwell, con sede en la Universidad de Alberta. El proyecto buscó catalogar diversas iniciativas de investigación en humanidades digitales alrededor del mundo, desde especialistas hasta entusiastas, mediante la donación de imágenes, textos, videos o cualquier otro medio que pudiera componer una muestra de la producción. En las definiciones dadas por los investigadores, las humanidades digitales fueron descritas como un estudio académico mediado por la computación, como interrogantes sobre el conocimiento en un área en que lo digital crea nuevas percepciones, como un área de intersección de varias áreas del conocimiento, como un área que carece de una definición más precisa, o incluso como una reinterpretación positivista de las humanidades a través de la computación.¹³

Hay aun más distinciones construidas dentro del campo del conocimiento para determinar qué se hace cuando se relaciona la investigación en humanidades con la computación. Específicamente en la historia, la historia digital se ha utilizado desde fines de la década de 1980 para definir la producción de historia en los medios digitales.¹⁴ Otra posible distinción es la de Richard Rogers, para quien existe una diferencia entre

los métodos digitales, oriundos del medio digital ya sea en datos o en formas, y los métodos digitalizados, que nacen en lo analógico y se vuelven digitales.¹⁵ Para él, esta sería la principal diferencia entre los métodos digitales y las humanidades digitales. Sin embargo, entendemos que tales distinciones solo fragmentan un debate que pertenece a todas las áreas del conocimiento, por lo que nos adherimos críticamente a las humanidades digitales como fundamento epistemológico, metodológico y hermenéutico.

La preocupación planteada aquí se relaciona con el uso de herramientas de investigación construidas a partir de métodos computacionales aplicados a documentos históricos. El trabajo consistió en utilizar buscadores de texto o desarrollarlos para documentos históricos. El resultado son nuevas formas de índices construidos a partir de los resultados de búsqueda. Si consideramos la historia de los libros, notamos que tecnologías como el índice general por capítulos, índice de nombres, notas al pie de página y bibliografía son invenciones modernas que se asocian al aumento de lectores, el número de escritores, la variedad de textos publicados y la mejora de las tecnologías editoriales.¹⁶ Por lo tanto, los nuevos índices representados en Crono.Data son, de alguna manera, solo parte de la historia de los libros relacionados con la historia social y la tecnología.

Desde el punto de vista de la relación de las mediaciones computacionales, Lev Manovich entiende que inauguraron un nuevo mundo, ya sea emulando elementos analógicos o construyendo objetos originales del mundo digital.¹⁷ Desde ese enfoque, el uso de la computadora modifica la forma en que interpretamos imágenes, textos, videos o cómo escribimos. Tal razonamiento puede derivarse de variaciones en las tecnologías computacionales, dado que la experiencia de un usuario en una computadora antigua con un símbolo del sistema y un monitor de tubo de imagen es totalmente diferente a la de una computadora con un procesador y una pantalla de última generación. La diferencia es aun mayor si pensamos en los anuncios realizados recientemente por empresas como Meta (antes Facebook) y Microsoft, que anunciaron una mayor inversión en el concepto de metaverso, es decir, un tipo de realidad virtual aumentada. En definitiva, la experiencia del usuario varía según la tecnología utilizada.

Considerando que el objeto de estudio en cuestión parte de documentos históricos digitalizados, es decir, son datos producidos en el mundo analógico y convertidos al digital, conviene entender un poco mejor qué opera cuando se realiza la transposición. Como escribió Manovich:

Es bastante apropiado (y más interesante) usar el término mapeo para describir lo que los nuevos medios hacen a los viejos medios. El *software* nos permite re-mapear objetos de medios antiguos en nuevas estructuras, transformando así los medios en lo que yo llamo “meta-medios”.¹⁸

Crono.Data, por tanto, es un metamedio construido a partir de medios tradicionales, que no se puede confundir con la lectura de un documento histórico *in loco*. La razón de la diferencia es que el investigador puede tener acceso a prácticamente todos los resultados del término de búsqueda, ya sea un concepto o un autor, pero no al documento completo. Esta opción se tomó para evitar conflictos con las leyes de derechos de autor, considerando que algunas colecciones de documentos históricos pueden tener derecho de uso de la digitalización o del propio documento. Replicarlos íntegramente equivaldría a copiar el trabajo desarrollado por colecciones digitales. Sin embargo, al extraer información del contenido de las colecciones, también se anota información sobre la ubicación de la colección, volumen, año y página, como un conjunto de metadatos para que el investigador pueda rastrear la información.

Además de la cuestión del contenido disponible, existe un extenso proceso desarrollado para recoger las informaciones, que transforma las imágenes escaneadas en archivos de base de datos para ser consultados y relacionados a través de Crono.Data. En el diagrama de flujo presentado a continuación (Figura 2) es posible tener una mejor comprensión de los procedimientos adoptados.

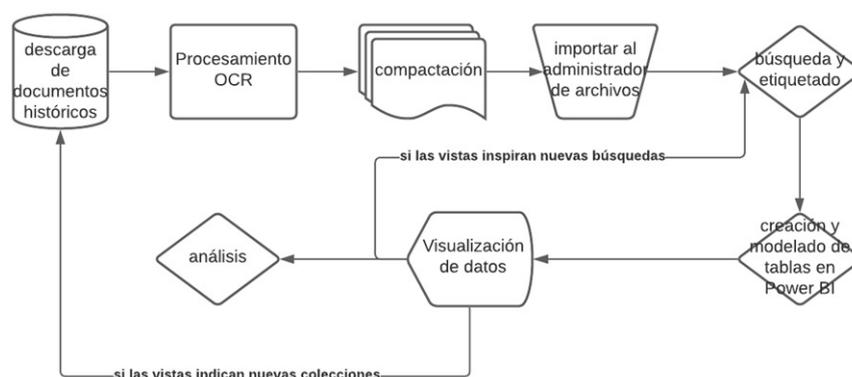


FIGURA 2

Diagrama de flujo de los pasos seguidos para preparar la base de datos y de investigación.

Elaboración propia.

El proceso de descarga de documentos históricos varía según la colección. Plataformas como la Biblioteca Brasiliana Guita y José Mindlin, la Biblioteca de Obras Raras de la UFRJ y la Biblioteca de la Academia de Historia Argentina ofrecen la descarga volumen por volumen con Reconocimiento Óptico de Caracteres (OCR)¹⁹ realizado. La colección del Instituto Histórico y Geográfico Brasileño proporciona volumen por volumen, pero sin OCR; esto significa que son solo imágenes guardadas en el formato Portable Document Format (.pdf). Para los archivos que son imágenes, considerando que el elemento escaneado es una fotografía, es necesario utilizar herramientas que lean los archivos y reconozcan los caracteres, transformándolos en un texto con capacidad de búsqueda. Existen numerosas herramientas con esta función; aquí se eligió Tesseract, considerando que es un *software* gratuito y permite guardar archivos en diferentes formatos, ya sea Portable Document Format (.pdf) o Comma-separated values (.csv), óptimo para base de datos.²⁰ Para colecciones como la Biblioteca Nacional, el método de descarga utilizado tiene otro enfoque, porque la plataforma solo pone a disposición página por página de cada libro o revista. Es decir, para un libro de 300 páginas, por ejemplo, se necesitan 300 descargas, lo que hace que el trabajo sea agotador. Para casos como este, se creó un robot programado en Python para automatizar el trabajo.²¹

La compresión es un proceso más simple, que se realiza solo cuando las imágenes tienen una resolución muy grande, lo que genera archivos pesados y ralentiza las búsquedas.²² Este es el caso de la colección de la Biblioteca Nacional, que, a pesar de dar acceso solo página por página para descargar, proporciona archivos de alta calidad. Es importante realizar el reconocimiento OCR antes de la compresión, porque influye en la posibilidad de reconocimiento de caracteres. Cuanto mejor sea la resolución del archivo, es más probable que los algoritmos de OCR reconozcan la mayoría de los caracteres. A continuación se muestra un ejemplo del volumen 6 de la *Revista do Instituto Histórico-Geográfico Brasileiro* (IHGB), de 1844, que contiene el famoso texto de Carl Friedrich Philipp von Martius, titulado “Cómo debería escribirse la historia de Brasil” (Figura 3). El texto ganó el concurso lanzado por el IHGB en 1840, que tenía como objetivo premiar el “Plan para escribir la Historia antigua y moderna de Brasil”.²³ En este mismo texto aparece por primera vez dentro de la Revista la palabra *historiografía*.

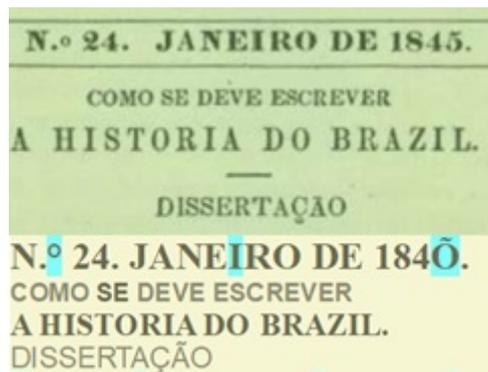


FIGURA 3

Comparación de una imagen digitalizada (superior), con su reconocimiento en OCR (inferior). Se percibe que el carácter 5 no fue reconocido.

Elaboración propia.

Es necesario considerar que no todos los caracteres se reconocen perfectamente, lo que impone limitaciones a la hora de la búsqueda. Para la palabra *historiografía*, por ejemplo, conviene hacer un tipo de búsqueda combinada, dividiendo el concepto para eludir posibles errores de reconocimiento, como “historio”, “torio” y “grafia”. A veces, los resultados arrojados repiten la ubicación de los términos, pero es preferible tener más datos para filtrar posteriormente, y no falta de informaciones que pueden ser relevantes. También es pertinente pensar en diferentes grafías, considerando que en textos antiguos palabras como “historiografía” se escriben con “ph”.

Para realizar la búsqueda es posible utilizar *softwares* lectores de Portable Document Format (.pdf) gratuitos así como Adobe Reader y Foxit Reader. Ambos en la configuración de búsqueda avanzada permiten seleccionar una carpeta determinada en la computadora y realizar búsquedas colectivas, además de guardar los resultados deseados. Otra opción es el uso de *software* especializado, como Atlas.ti, Nvivo y Maxqda; sin embargo, dichas herramientas deben comprarse al fabricante. La tercera opción es la más compleja, pero la más poderosa para el análisis de datos, y consiste en utilizar lenguajes de programación como Python. Su uso requiere de otras herramientas proporcionadas por el fabricante, que son gratuitas y cuentan con una amplia documentación dada la gran comunidad activa de usuarios.²⁴

Existen numerosos lenguajes de programación y formas de enfoques dentro de ellos que les permiten lograr el mismo resultado. A pesar de usar Python, se observa que es posible conseguir resultados similares con SQL (Structured Query Language), por ejemplo. Como escribe Robert W. Sebesta, “El lenguaje en el que ellos [los programadores] desarrollan *software* impone restricciones sobre los tipos de estructuras de control, estructuras de datos y abstracciones que pueden usar”.²⁵ Python ha sido uno de los lenguajes más utilizados en el mundo, según los datos de PYPL, no solo porque es de código abierto, porque tiene una comunidad activa, o por las numerosas bibliotecas de código libre, sino también por su facilidad de aprendizaje, muy parecido al de la escritura de la propia lengua inglesa.²⁶

Para buscar archivos históricos usando Python es necesario guardarlos en formato Comma-separated values (.csv) o convertirlos desde otro archivo.²⁷ La gran ventaja aquí es extraer información de documentos históricos y reubicarla en otros documentos. Es precisamente este trabajo el que permite la correlación de información de forma interactiva en Crono.Data. Crono.Data es solo un relator visual de la información contenida en las tablas construidas, que contiene datos sobre la página, el volumen y la ubicación de la colección. Elegimos utilizar la plataforma de Repositório de Dados de Pesquisa da Unicamp para que dichos metadatos estén disponibles para su descarga.²⁸ A continuación (Figura 4) se muestra un pequeño ejemplo de cómo se construyen las tablas.

trecho	nome do arquivo	Ano	palavra buscada	página
costuma ter	extasi. Extemporanean Dicionário da lingua brasileira luiz da maria pinto	1832	extemporâneo	497
ou feito cx-	temporaneamente , de re Dicio R. Bluteau reformado por Antonio M. Silva v.1	1789	extemporâneo	611
orar Vieira „	cçmpuferãó— o hymnc Dicio R. Bluteau reformado por Antonio M. Silva v.1	1789	extemporâneo	611
enganadas	SEÇÃO VII — CRIADOR DO Dicio Analgico Da Lingua Portu Carlos Spitzer	1936	ressentimento	20
relembra, memó-	ria do beneficio, d. Dicio Analgico Da Lingua Portu Carlos Spitzer	1936	ressentimento	165
Seção VII —	CRIADOR DO PENSAMEN Dicio Analgico Da Lingua Portu Carlos Spitzer	1936	ressentimento	167
con-	flito, colisão, luta, antagonismo, Dicio Analgico Da Lingua Portu Carlos Spitzer	1936	ressentimento	231

FIGURA 4

Tabla con datos sobre la búsqueda de conceptos utilizados por Nietzsche.

Elaboración propia.

La parte visual se basa en Microsoft Power Bi, que es una herramienta para crear gráficos interactivos. Hay varias herramientas de este tipo disponibles, como Tableau y Google Data Studio. Tienen funciones gratuitas y de pago. Se eligió Power Bi porque ofrece más funciones de forma gratuita. Estos elementos visuales interactivos de tablas también se pueden construir en programas como Excel o LibreOffice Calc; sin embargo, en estas condiciones se requieren conocimientos avanzados en el lenguaje Visual Basic. En herramientas como Power Bi, Data Studio y Tableau, basta conocer las interfaces de los programas, los cuales son muy intuitivos y cuentan con numerosos videotutoriales en redes sociales. Los conceptos para relacionar datos provienen del área de la computación denominada modelado de datos, es decir, cómo pensar y desarrollar datos siguiendo una arquitectura estructurada y eficiente. Los conceptos de entidades y relaciones son pertinentes aquí. En la bibliografía sobre modelado de datos hay varias subdistinciones de los tipos de entidades y relaciones que se utilizan para bases de datos más complejas y extensas.²⁹

Para entender cómo funciona Crono.Data, basta saber que las entidades son las columnas de la tabla anterior, como extracto, nombre de archivo, palabra buscada, año y página. Ahora bien, si tengo otra tabla con otra información será posible establecer la relación a través de alguna entidad que tenga el mismo nombre, ya sea el año, el número de página o el nombre del archivo. Por tanto, basta utilizar una entidad estándar en las tablas, como el año o el nombre de la colección, para cruzar los datos y generar nuevas formas de representación.

Se observa, así, que se necesitan numerosos tipos de conocimientos informáticos para pensar y operar los datos que se utilizarán en la investigación de la historia. Hay contacto con opciones y enfoques tecnológicos que modifican no solo la experiencia del investigador, sino el producto final que se analiza en cuanto datos de investigación. Ahora, está claro que hay un proceso de construcción de índices y gestión de documentos históricos, hasta que estén disponibles para un análisis hermenéutico de los datos.

CANTIDAD Y CALIDAD

Aunque las investigaciones referidas a humanidades y computación son más frecuentes a partir del siglo XXI, algunas de las críticas contemporáneas tienen variaciones más antiguas a través de otros debates, como los métodos cuantitativos. En 1968, cuando Emmanuel Le Roy Ladurie dijo que “El historiador del mañana será programador o no”,³⁰ había un cierto entusiasmo por parte de algunos historiadores por los métodos cuantitativos, seriados, estadísticos y matemáticos, realzado por la perspectiva de la computación.

En ese momento, ya era posible notar las tensiones entre lo cuantitativo y lo cualitativo en el texto clásico *History from Below*, publicado en 1966 por Edward Palmer Thompson. El autor pensó en métodos de estudio para el ciudadano común, sus hábitos, ocio, valores familiares, en definitiva, aspectos culturales que los problemas derivados de la economía y sociología aplicadas a la historia parecían simplificar en frases extraídas de datos.³¹ Entre los textos más conocidos por valorar los cortes cualitativos se encuentran *Las palabras y las cosas*, *La arqueología del saber* e *Historia de la sexualidad* de Michel Foucault. Años más tarde, François Dosse, en su crítica del movimiento de los *Annales*, enfatizó las tensiones entre lo cuantitativo y lo cualitativo,

entre la continuidad y la discontinuidad, entre lo serial y lo disruptivo, entre lo específico y lo general en la producción de la historia.³² Si bien libró a Le Roy Ladurie de las duras críticas que hacía al movimiento de los *Annales*, debido a las constantes innovaciones teóricas del autor, no dejó de categorizar la fiebre de la cuantificación como una respuesta a la fragmentación de las narrativas, que cobró fuerza a partir de mayo de 1968.³³

Entre los años sesenta y setenta, la producción de la historia vivió lo que Pierre Nora denominó el “periodo historiográfico”;³⁴ considerando el giro crítico sobre el oficio mismo: se plantearon interrogantes sobre métodos y teorías, y se desnaturalizó tanto el lugar como la escritura del oficio del historiador, en la célebre forma de Michel de Certeau. Cualquier forma de escritura está bajo el juicio crítico de un historiador o historiadora, dados los puntos de inflexión en los debates de las últimas décadas, como el giro lingüístico, el giro representacional, el giro ético, el giro de género o, más recientemente, el giro digital. Buscaré otra formulación, aunque existen buenos ejemplos para la defensa del carácter cuantitativo en la producción de la historia, como lo hace Antoine Proust al cuestionar el fundamento del oficio en relación con las ciencias sociales y la economía, escribiendo que los “los historiadores entenderían mejor con los métodos cuantitativos —muchas veces indispensables— si prestaran más atención a desmitificar cifras y cálculos”.³⁵

Es necesario distinguir las formas cuantitativas operadas en el siglo XX por los recursos utilizados en la Crono.Data. Si en el apogeo de la historia de las mentalidades, en la historia económica o en la historia geológica se usaron cálculos para cuantificar la producción agrícola, el número de trabajadores o las variaciones de temperatura, las matemáticas utilizadas aquí son solo una pequeña parte del método. Por supuesto, en Crono.Data existen cálculos matemáticos que cuentan el número total de veces que aparece una palabra o autor, pero la expresión se basa en la terminología que se quiere buscar en libros o revistas, que necesitan ser analizado *in loco*. Por ejemplo, si seleccionamos la palabra *historiografía* en la colección de la *Revista do IHGB*, nos encontramos ante un gráfico (Figura 5) que muestra tímidas apariciones en el siglo XIX, fluctuaciones a mediados del siglo XX y una curva ascendente después de la década de 1970.

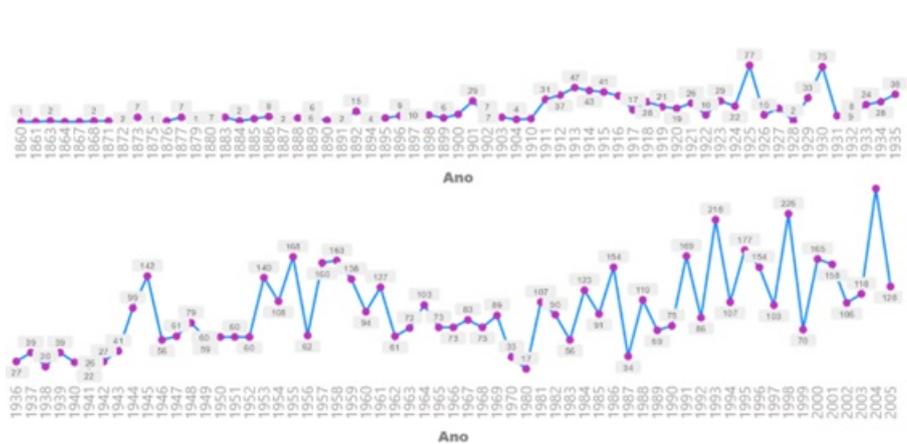


FIGURA 5

Gráfico de 1860 a 2005 con la aparición de la palabra "historiografía" en la *Revista do IHGB*.

Elaboración propia.

El gráfico permite interpretaciones según la erudición de cada historiador o historiadora, observando los fenómenos a la luz de la historia de los conceptos, como en Reinhart Koselleck, que nos permite imaginar tendencias, diacronismos y sincronismos, o mediante hechos históricos, como la caída de la Bolsa o la Segunda Guerra Mundial, lo que puede imponer nuevas lógicas en la publicación de libros y revistas según la editorial o el país. Sin embargo, si utilizamos métodos computacionales para, además de cuantificar, leer la cita, se abren otras posibilidades interpretativas. Por ejemplo, durante los primeros cien años de la *Revista do IHGB*, la palabra *historiografía* se articuló de diferentes maneras. En el año 1844 fue utilizada por Von Martius para

distinguir una historia pragmática sobre el pasado de una historia filosófica, es decir, especulativa. Llama la atención en el mismo volumen el uso de la palabra *historiografía* por Adolfo de Varnhagen, quien conservó la raíz de la palabra en portugués hasta finales del siglo XIX. En este periodo, el historiógrafo cambió el significado referido a alguien que estaba autorizado por los reyes a escribir historia, a alguien cuya escritura era autorizada por las cualidades de la erudición, el “grande-*historiographo*”. Recién en 1899 se utilizó la palabra *historiographia* con uno de los significados que tiene hoy, referida a un conjunto de textos producidos sobre un determinado tema.

Por supuesto, a partir de estos análisis semánticos, también se pueden observar otros innumerables aspectos, dependiendo del problema planteado por el historiador. Si se nota en el gráfico el crecimiento de la palabra *historiografía* en el año 1925, se debe a una revisión de Max Fleiuss del libro del historiador argentino Rómulo D. Carbia, titulado *Historia de la historiografía argentina*. Este es solo un episodio de las relaciones desarrolladas entre historiadores brasileños y argentinos, quienes durante la primera mitad del siglo XX trabajaron juntos para publicar libros, organizar conferencias, fundar instituciones de investigación, políticas de financiamiento, intercambios e incluso acuerdos diplomáticos. La interacción de la constelación del autor también se puede ver en el Crono.Data (Figura 6).

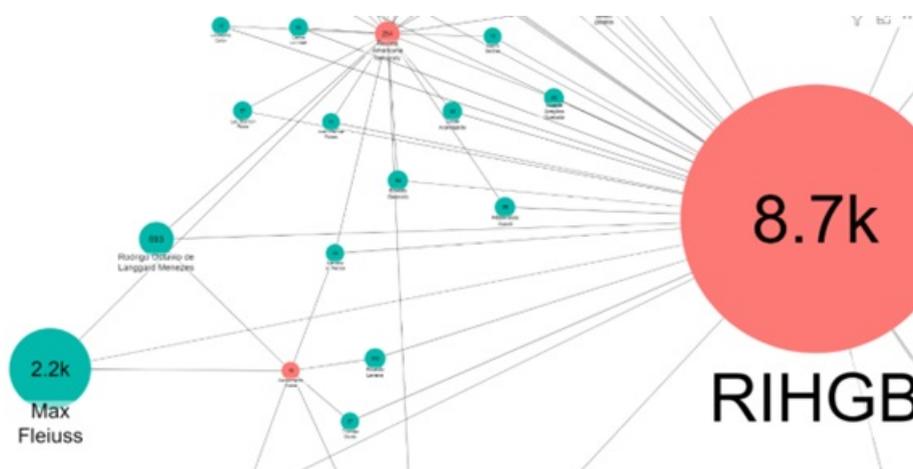


FIGURA 6

Gráfico elaborado a partir de grafos que relacionan a los autores (bolas azules) con colecciones (bolas naranjas). El valor dentro del círculo representa el número de citas.

Elaboración propia.

Algunos trabajos en Brasil han comenzado recientemente a estudiar esta interacción intelectual entre Brasil y Argentina en las primeras décadas del siglo XX, destacando dos aspectos, uno más centrado en las relaciones internacionales, en la obra de Raquel Paz, y otro de carácter historiográfico, como en las obras de Gabriela Oliveira, Ana Paula Barcelos, Maria Pascual Guimarães.³⁶ Son trabajos enfocados en situaciones específicas, como los acuerdos bilaterales entre los ministerios de Relaciones Exteriores de Brasil y Argentina, la relación entre el IHGB y la Unión Panamericana, y las relaciones entre intelectuales como Max Fleiuss y Ricardo Levene. Sin embargo, el estudio de las redes intelectuales aumentadas por datos cuantitativos y cualitativos obtenidos a través de métodos digitales nos ha permitido medir el nivel y la extensión de los intercambios. Se mapearon los intelectuales que tenían intercambios entre Brasil y Argentina y, a través de los datos prosopográficos contenidos en las *Revista do IHGB* y en el *Boletim da Junta de História e Numismática Americana*,³⁷ fue posible trazar cómo se relacionaba la red.

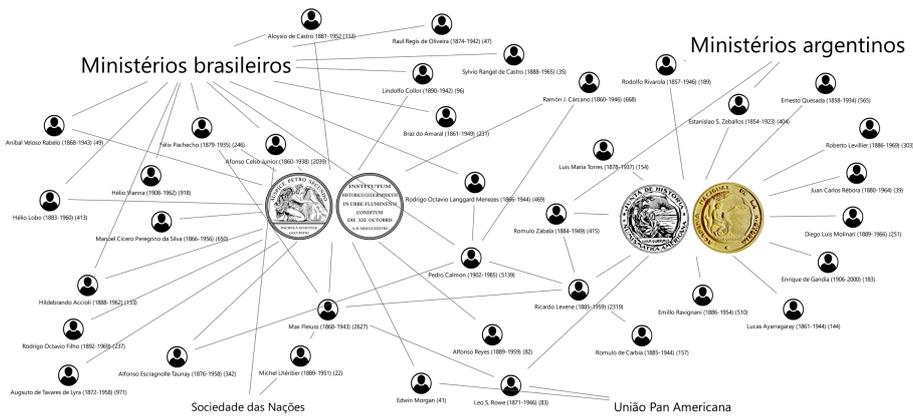


FIGURA 7
 Red mapeada a partir del *Boletim da Junta de História e Numismática Americana* y la *Revista do IHGB*, entre los años 1920 a 1940.
 Elaboración propia.

En la Figura 7 se puede ver que Afonso Celso Junior, Rodrigo Octavio Langgard, Max Fleiuss y Pedro Calmon fueron los principales articuladores del lado brasileño. Por el lado argentino, Ramón Cárcano, Rodolfo Rivarola y Ricardo Levene.³⁸ También fue posible notar distinciones en la forma de actuar. Algunos intelectuales se involucraron más en trabajos sobre la escritura de la historia, como en el I Congreso Internacional de Historia de América, realizado en Río de Janeiro en 1922, en el que Fleiuss fue uno de los principales articuladores; en la Biblioteca de Autores Traducidos al castellano y al portugués (1935), en la que Calmon y Levene trabajaron juntos; y en el II Congreso Internacional de Historia de América, realizado en Buenos Aires bajo la mirada de Levene. Otros intelectuales como Afonso Celso, Rodrigo Octavio, Ramón Cárcano y Rodolfo Rivarola estuvieron más vinculados a la política internacional, forjando acuerdos que propiciaron políticas culturales para ambos países que culminaron en el Instituto Brasileiro-Argentino de Cultura y el Instituto Argentino-Brasileño de Cultura.

Lista de Códigos	Aug.	Man.	Max.	Pedr.	Alfo.	Rod.	Rod.	Afonso Ce.	Rod.	Enri.	Esta.	Luc.	Ram.	Ro.	Rob.	Erne.	Emil.	Rica.
Augusto de Tavares de Lyra (1872-15)		118	207	109	22	56	64	178		1	5	6	16					4
Manoel Cícero Peregrino da Silva (1866-1943)	118		203	44	8	23	77	229	2	1	1	2	9	1				3
Max Fleiuss (1868-1943)	207	203		174	28	80	141	513	11	7	3	12	23	3	11	5	3	53
Pedro Calmon (1902-1985)	109	44	174		15	57	42	72	7	3	2	1	13	9	13	1	8	44
Afonso Escagnolle Taunay (1876-1955)	22	8	28	15		7	2	17					5	1				5
Rodrigo Octavio Filho (1892-1969)	56	23	80	57	7		27	48					7					2
Rodrigo Octavio Langgard Menezes (1866-1944)	64	77	141	42	2	27		134	7	2	1	2	18	1	1	2		12
Afonso Celso Junior (1860-1938)	178	229	513	72	17	48	134		5	3	3	10	37	2	1	9	2	15
Rodolfo Rivarola (1857-1946)		2	11	7	1		7	5			1	31	47	34	1	10	11	43
Enrique de Gandia (1906-2000)	1	1	7	3		1	2	3	2		1	6	6	25	3	2	3	35
Estanislao S. Zeballos (1854-1923)	5	1	3	2			1	3	1	1			2	16	3		43	2
Lucas Ayarragaray (1861-1944)	6	2	12	1		2	10	31	6	2			47	32		10	3	44
Ramón J. Cárcano (1860-1946)	16	9	23	13	5	7	18	37	47	6	16	47	62	3	36	20	171	171
Romulo Zabala (1884-1949)		1	3	9	1		1	2	34	25	3	32	62		5	11	37	158
Roberto Levíer (1886-1969)			11	13			1	1	1	1			3	5		1	7	29
Ernesto Quesada (1858-1934)	4	3	5	1		1	2	9	10	2	43	10	36	11	1		1	34
Emilio Ravignani (1880-1954)			3	8			2	11	3	2	3	20	37	7	1			120
Ricardo Levene (1885-1959)	4	3	53	44	5	2	12	15	43	35	8	44	171	158	29	34	120	120

FIGURA 8
 Cruce entre intelectuales brasileños y argentinos según las colecciones del I y II Congreso Internacional de Historia de América, *Boletim da Junta/Academia* y *Revista do IHGB*. De 1920 a 1940.
 Elaboración propia.

En la Figura 8 se representa la medición de la relación entre los principales nombres de la red intelectual que mapeamos. La relación se construye a partir de la intersección entre los intelectuales catalogados, ya que un autor es citado junto a otro. La relación puede significar la simple citación de un intelectual, la citación de intelectuales por un tercero o la participación en las reuniones de la Junta de Historia y del IHGB, las cuales fueron registradas en actas y publicadas junto a los diarios. Por el lado brasileño hubo constantes intercambios entre Tavares Lyra, Manoriel Cícero Peregrino, Max Fleiuss, Pedro Calmon, Rodrigo Octavio

Langgard y Afonso Celso. Del lado argentino, Levene fue el principal punto de articulación entre la Junta y otros intelectuales como Lucas Ayarragaray, Ramón Cárcano, Rómulo Zabala y Emilio Rivignani.

Ricardo Levene fue presidente de la Junta de Historia y Numismática entre los años 1927-1931, 1934-1953 y 1955-59. Según los estudios de Fernando Devoto y Pagani Nora, la Junta reunió a historiadores de distintas generaciones, posiciones teóricas y políticas, pero en la figura de Levene logró una cierta coalición a favor de movimientos institucionales que permitieron la profesionalización y expansión del campo historiográfico en el país.³⁹ Por ese motivo, varios historiadores de la llamada Nueva Escuela Argentina, como Emilio Ravignan, Diego Luis Molinari y Luis M. Torres también frecuentaban la Junta.

Entre los historiadores brasileños y argentinos se pueden observar diferencias en los usos de conceptos como Panamérica, América Latina, América del Sur o Hispanoamérica. La palabra Panamérica fue mucho más común entre los intelectuales brasileños, mientras que Hispanoamérica fue utilizada casi exclusivamente por los intelectuales argentinos. Las excepciones están en el I y II Congreso Internacional de Historia de América, en 1922 y 1937, que contaron con numerosos intelectuales de países americanos, lo que significó diferentes posiciones políticas en relación con los Estados Unidos. En la definición de Alexandra Pita, *panamericanismo* “remite a todos los países del continente americano, mientras que los otros como Latino, Ibero e Hispano excluyen a Estados Unidos”.⁴⁰ Variando según cada intelectual argentino, también se excluyó a Brasil. Si bien Brasil y Argentina intentaron aproximaciones durante las primeras décadas del siglo XX,⁴¹ hubo disputas por la hegemonía de América del Sur entre ambos países.⁴² Los intereses de Estados Unidos en la región, explícitos desde la I Conferencia Americana de Washington en 1890, introdujeron nuevas variables al debate.⁴³

Revista IHGB					
Conceitos	Ibero-América	América do Sul	América Latina	Hispano-América	Pan-América
Ibero-América	0	1	0	0	0
América do Sul	1	0	6	7	11
América Latina	0	6	0	3	5
Hispano-América	0	7	3	0	3
Pan-América	0	11	5	3	0
Boletim Junta					
Conceitos	Ibero-América	América do Sul	América Latina	Hispano-América	Pan-América
Ibero-América	0	0	0	1	0
América do Sul	0	0	0	0	0
América Latina	0	0	0	2	2
Hispano-América	1	0	2	0	1
Pan-América	0	0	2	1	0

FIGURA 9

Comparación de la asociación entre conceptos, de 1924 a 1950.

Elaboración propia.

En la Figura 9 es posible percibir la relación entre los conceptos de Iberoamérica, América del Sur, América Latina, Hispanoamérica y Panamérica, empleando el mismo criterio de intersección utilizado para los intelectuales. Se puede ver que la terminología Hispanoamérica nunca se relacionó con Panamérica en el *Boletim da Junta*, mientras que en la *Revista do IHGB* era muy común utilizar los significados de América del Sur y América Latina como sinónimos de Panamérica.

Cabe mencionar que la red que mapeamos tenía operaciones constantes en otros países, como Uruguay, Chile y México. Y del lado brasileño, también hubo acción con la Sociedad de las Naciones, en Europa, y sus derivaciones institucionales, especialmente en los debates sobre cooperación intelectual.

CONSIDERACIONES FINALES

Los análisis realizados aquí están todavía en pleno desarrollo de investigación. Crono.Data sigue preparándose y pronto contendrá nuevas colecciones que aún se están indexando. También se estudia la posibilidad de albergar datos de otras encuestas digitales, que involucren otros recortes y tipos de documentación, como Twitter. Las tres formas de búsquedas en documentos históricos que se presentan son una demostración del carácter experimental que estamos llevando a cabo. Sin embargo, a pesar de esta condición experimental, a lo largo del texto buscamos hacer un uso restringido de algoritmos para que prevalezca la habilidad hermenéutica del historiador.

BIBLIOGRAFIA

- Cervo, Amado Luis y Wolfgang Dopcke. *Relações internacionais dos países americanos: Vertentes da história*. Brasília: UNB, 1994.
- Chartier, Roger. *A aventura do livro*. São Paulo: Unesp, 1998.
- De Castro, Leandro N. y Daniel G. Ferrati. *Introdução à mineração de dados: Conceitos básicos, algoritmos e aplicações*. São Paulo: Editora Saraiva, 2016.
- De Sordi, José Osvaldo. *Modelagem de dados-estudos de casos abrangentes da concepção lógica à implementação*. São Paulo: Saraiva Educação SA, 2019.
- Dosse, François. "Uma história serial", en *A história em migalhas*. Campinas: Ed da Unicamp, 1994.
- Devoto, Fernando y Nora Pagano. *Historia de la historiografía argentina*. Buenos Aires: Sudamericana, 2009.
- Dosse, François. *A história à prova do seu tempo*. São Paulo: Editora Unesp, 2001.
- Guimarães, Lucia Maria Paschoal. *Da escola palatina ao silogeu: Instituto Histórico e Geográfico Brasileiro*. Rio de Janeiro: Museu da República, 2007.
- Kantardzic, Mehmed. *Data mining. Concepts, models, methods, and algorithms*. New Jersey: IEEE Press, 2003.
- Le Roy Ladurie, Emmanuel. *Le territoire de l'historien*. Paris: Gallimard, 1977.
- Manovich, Lev. "Visualização de dados como uma nova abstração e anti-sublime", en *Derivas: Cartografias do ciberespaço*, ed. por Lúcia Leão (São Paulo: Sesc, 2004).
- Manovich, Lev. *The language of new media*. Massachusetts: MIT Press, 2002.
- Nora, Pierre. "Entre mémoire et histoire. La problématique des lieux: Les lieux de mémoire", en *La République*, ed. por Pierre Nora. Paris: Gallimard, 1984.
- Proust, Antoine. *Doze lições sobre História*. Belo Horizonte: Autentica, 2008.
- Revista del Instituto Histórico e Geográfico Brasileiro*, Rio de Janeiro, 2 (1840): 642.
- Sabesta, Robert W. *Conceitos de linguagens de programação*. Porto Alegre: Bookman, 2011.
- Santos, Raquel Paz. *Um novo olhar sobre o país vizinho*. Rio de Janeiro: Multifoco, 2012.
- Santos, Luis Carlos Villafañe. *O Brasil entre a Europa e a América: o Império e o interamericanismo (do Congresso do Panamá à Conferência de Washington)*. São Paulo: Unesp, 2004.
- Thompson, Edward Palmer. "A história vista de baixo", en *A peculiaridade dos ingleses e outros ensaios*, ed. por Antonio Luigi Negro. Campinas: Editora Unicamp, 2012.

DIGITAL

- Barcelos, Ana Paula. "Diálogos sobre a escrita da história. Ibero-americanismo, catolicismo, (des)qualificação e alteridade no Brasil e na Argentina". Tesis, ICHF, UFF, 2011.

- Busa, Roberto et al. "Index Thomisticus". *Corpus Thomisticus*. Acceso el 31 de diciembre de 2021. <https://www.corpusthomisticum.org/it/index.age>
- Timm, Maria Isabel et al. "A virada computacional da filosofia e sua influência na pesquisa educacional". *Ciências & Cognição* 11 (2007).
- Allington, Danielle et al., "Neoliberal tools (and archives): A political history of digital humanities". *Los Angeles Review of Books*. Acceso el 12 de mayo de 2021. <https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities>
- THATCamp. "Manifiesto das digital humanities". *Hypotheses. The Humanities and Technology camp. 2012*. Acceso el 13 de enero de 2020. <https://tcp.hypotheses.org/497>
- Gold, Matthew K. "Day of DH: Defining the Digital Humanities", en *Debates in the Digital Humanities*, ed. por Matthew K. Gold. London: University of Minnesota Press, 2012.
- Ayers, Edward L. "The pasts and futures of digital history". *History News* 56, n° 4 (2001): 5-9.
- Müller, Bertrand. "Lucien Febvre et Henri Berr: De sa synthèse à l'histoire-problème", *Revue de Synthèse* 117 n.° 1-2 (1996): 39-59. doi: 10.1007/BF03181302
- Pita, Alexandra. "Pan-americanismo y nación", *Anuario IEHS* 32 n.° 1 (2017).
- Rogers, Richard. *Digital methods*. Cambridge, Massachusetts: Mit Press, 2013.
- Silva, Gabriela Correa. "Dos passados heterogêneos ao mosaico continental: pan-americanismo e operação historiográfica no IHGB republicano (1889-1933)". Tesis, IFCH, UFRGS, 2019.
- Stamatatos, Efstathios. A survey of modern authorship attribution methods. *JASIST* 60, n.° 3 (2009): 538:556.

SITES

- Python Software Foundation. "Python 3.10.1 documentation". Acceso el 6 de noviembre de 2021. <https://docs.python.org/3/>
- Tesseract. *GitHub*. Acceso el 5 de agosto de 2019. <https://github.com/tesseract-ocr/tesseract>
- Comprimir. *Ilovepdf*. Acceso el 25 de mayo de 2020. https://www.ilovepdf.com/pt/comprimir_pdf
- Comprimir. *Smallpdf*. Acceso el 25 de mayo de 2020. <https://smallpdf.com/pt/comprimir-pdf>
- Pdf-Csv. *Convertio*. Acceso el 19 de noviembre de 2021. <https://convertio.co/pt/pdf-csv/>
- Pdf-to-csv. *CDKM*. Acceso el 19 de noviembre de 2021. <https://cdkm.com/pt/pdf-to-csv>
- Rota, Alesson R. "History mining". *Redu-Unicamp*. Acceso el 19 de noviembre de 2021. <https://doi.org/10.25824/redu/CRM7ZU>
- Data Science Journal. Volume 1 and 2. *Codata*. Acceso el 18 de diciembre de 2021. <https://datascience.codata.org/9/volume/1/issue/1/>
- Rota, Alesson R. Crono.Data. História da História. Acceso el 12 de diciembre de 2021. <http://historiadahistoria.digital/>
- Oracle. What is data science. *Oracle*. Acceso el 12 de diciembre de 2021. <https://www.oracle.com/br/data-science/what-is-data-science/>
- PYPL PopularitY of Programming Language. *GitHub*. Acceso el 25 de mayo de 2020, <https://pypl.github.io/PYPL.html>
- Rackwitz, Christoph. Tesseract ocr array of images to text csv. *Stackoverflow*. Acceso el 27 de noviembre de 2021, <https://stackoverflow.com/questions/69239883/tesseract-ocr-array-of-images-to-text-csv>

Códigos⁴⁴

Algoritmo para convertir la imagen a OCR y guardarla en .CSV 45

```
import pytesseract
from PIL import Image # pip install Pillow
# comando para importar Tesseract
pytesseract.pytesseract.tesseract_cmd = r'/usr/local/bin/tesseract'
# comando para iniciar y guardar en la carpeta indicada -> '/usr/local/bin/tesseract'
list_with_many_images = [
    "./imagetocsv/img/Image_1.jpg",
    "./imagetocsv/img/Image_2.jpg",
    "./imagetocsv/img/Image_3.jpg", ] # crea una función que devuelve el texto def image_to_str(path):
    """ devolver una string (texto) de la imagen """
    return pytesseract.image_to_string(Image.open(path))
# convertir y guardar a CSV como queramos
with open("images_content.csv", "w+", encoding="utf-8", newline="") as file:
    file.write("Image_id,Image_Name,Image_Text,Date")
for image_path in list_with_many_images:
    text = image_to_str(image_path)
    line = f"{image_path}, {text}" file.write(line)
```

Algoritmo para descargar las colecciones de la Biblioteca Nacional de Brasil 46

```
from requests_html import HTMLSession
DEFAULT_USER_AGENT = 'Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_6) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/85.0.4183.102 Safari/537.36'
session = HTMLSession();
headers = {'User-Agent': DEFAULT_USER_AGENT}
DOMAIN = 'http://memoria.bn.br/docreader/'
print("Fetching images...")
# variables creadas para recibir datos del sitio indicado.
def download_image(url, page):
    response = session.get(url, headers=headers)
    response.html.render()
    img = response.html.find('img');
    print(img)
    with open('./revista-americana/1919/011-012/page-' + str(page).zfill(3) + '.jpg', 'wb') as w:
        w.write(response.content)
# modificar el link "/revista-americana/1919/011-012/page-" descargar volúmenes de otros años
list = [ ]
start = 11627-1;
for index in list: range(1, 161):
# los valores dentro del rango son el rango de páginas que se descargarán
page = start + index;
r =
```

```

session.get('http://memoria.bn.br/docreader/DocReader.aspx?bib=052558&pagfis=' + str(page),
headers=headers)
r.html.render(sleep=16)
id = r.html.find('#DocumentoImg', first=True);
print(id.attrs.get('src'))
download_image(DOMAIN + id.attrs.get('src'), index);

```

NOTAS

1. Todas las citas en idiomas extranjeros han sido traducidas al español por el autor.
2. Oracle. What is data science. *Oracle*. Acceso el 12 de diciembre de 2021. <https://www.oracle.com/br/data-science/what-is-data-science/>
3. *Data Science Journal*. Volume 1 and 2. Codata. En <https://datascience.codata.org/9/volume/1/issue/1/>
4. Leandro de Castro, Daniel G. Ferrari, *Introdução à mineração de dados: Conceitos básicos, algoritmos e aplicações* (São Paulo: Editora Saraiva, 2016).
5. Mehmed Kantardzic, *DATA MINING Concepts, Models, Methods, and Algorithms* (New Jersey: IEEE Press, 2003).
6. Alesson R. Rota, "Crono.Data. História da História". Acceso el 12 de diciembre de 2021. <http://www.historiadahistoria.digital/>
7. Roberto Busa et al., "Index Thomisticus". *Corpus Thomisticus*. Acceso el 31 de diciembre de 2021. <https://www.corpusthomisticum.org/it/index.age>
8. Efstathios Stamatatos, "A survey of modern authorship attribution methods". *JASIST* 60, n.º 3 (2009): 538-556.
9. Leslie Burkholder, *New Journal: Philosophy & Computing*, citado por Maria Isabel Timm et al., "A virada computacional da filosofia e sua influência na pesquisa educacional". *Ciências & Cognição* 11 (2007).
10. En el caso específico de Moretti, el tema no es tanto computacional, sino un conjunto de datos y estadísticas que complementan sus análisis literarios.
11. Danielle Allington et al., "Neoliberal Tools (and Archives): A Political History of Digital Humanities", en *Los Angeles Review of Books*. Acceso el 12 de mayo de 2021. <https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities>
12. THATCamp. Manifiesto das digital humanities. *Hypotheses. The Humanities and Technology Camp. 2012*. Acceso el 13 de enero de 2021. <https://tcp.hypotheses.org/497>
13. Matthew K. Gold, "Day of DH: Defining the Digital Humanities", en *Debates in the Digital Humanities*, ed. por Matthew K. Gold (London: University of Minnesota Press, 2012).
14. Edward L. Ayers, "The pasts and futures of digital history". *History News* 56, n.º 4 (2001):5.
15. Richard Rogers, *Digital Methods* (Cambridge, Massachusetts: Mit Press, 2013), 43.
16. Roger Chartier, *A aventura do livro* (São Paulo: Unesp, 1998).
17. Lev Manovich, *The language of new media* (Massachusetts: MIT Press, 2002).
18. Lev Manovich, "Visualização de dados como uma nova abstração e anti-sublime", en *Derivas: Cartografias do ciberespaço*, ed. por Lúcia Leão (São Paulo: Sesc, 2004).
19. OCR es la tecnología empleada para reconocer caracteres en imágenes y convertirlos en texto en formato digital. Actualmente existen otras tecnologías en uso, como el reconocimiento inteligente de caracteres (ICR), que involucra inteligencia artificial, y el reconocimiento de palabras escritas a mano (IHR).
20. Puede acceder a los documentos del programa, tutoriales y descarga en: Tesseract. *Github*. En <https://github.com/tesseract-ocr/tesseract>
21. El código del robot se encuentra en el Apéndice.
22. Para disminuir la resolución de las imágenes existen sitios web que hacen el trabajo gratuitamente, como Comprimir. [ilovepdf](https://www.ilovepdf.com/es/comprimir_pdf), en https://www.ilovepdf.com/es/comprimir_pdf; Smallpdf, en <https://smallpdf.com/pt/comprimir-pdf>
23. *Revista del Instituto Histórico e Geográfico Brasileiro*, Rio de Janeiro, 2 (1840):642.
24. Python Software Foundation, "Python 3.10.1 documentation". Acceso el 6 de noviembre de 2021. <https://docs.python.org/3/>
25. Robert W. Sabesta, *Conceitos de linguagens de programação* (Porto Alegre: Bookman, 2011), 20.
26. PYPL PopularitY of Programming Language. *Github*. Acceso el 25 de mayo de 2020. <https://pypl.github.io/PYPL.html>
27. Para convertir es posible utilizar plataformas online gratuitas, como Pdf-Csv. *Convertio*. en <https://convertio.co/pt/pdf-csv/>; Pdf-to-csv. *CDKM*. en <https://cdkm.com/pt/pdf-to-csv>. Incluimos un apéndice con el código para guardar los archivos de OCR directamente en .CSV, a través de Python y de Tesseract.
28. Alesson R. Rota, "History mining". *Redu-Unicamp*. En <https://doi.org/10.25824/redu/CRM7ZU>

29. José Osvaldo de Sordi, *Modelagem de dados-estudos de casos abrangentes da concepção lógica à implementação* (São Paulo: Saraiva Educação SA, 2019).
30. El texto se publicó originalmente en la revista *Le Nouvel Observateur* bajo el título “La fin des érudits”, y luego se volvió a publicar en el libro *Le territoire de l'historien*, bajo el título “L'historien et l'ordinateur”. Emmanuel Le Roy Ladurie. *Le territoire de l'historien* (Paris: Gallimard, 1977).
31. Edward Palmer Thompson, “A história vista de baixo”, en *A peculiaridade dos ingleses e outros ensaios*, ed. por Antonio Luigi Negro (Campinas: Editora Unicamp, 2012), 195-198.
32. François Dosse, “Uma história serial”, en *A história em migalhas* (Campinas: Ed da Unicamp, 1994).
33. François Dosse, *A história à prova do seu tempo* (São Paulo: editora Unesp, 2001).
34. Pierre Nora, “Entre mémoire et histoire. La problématique des lieux: Les lieux de mémoire”, en *La République*, ed. por Pierre Nora (Paris: Gallimard, 1984).
35. Antoine Proust, *Doze lições sobre História* (Belo Horizonte: Autentica, 2008), 63.
36. Raquel Paz Santos, “Um novo olhar sobre o país vizinho (Rio de Janeiro: Multifoco, 2012). Gabriela Correa Silva, Dos passados heterogêneos ao mosaico continental: pan-americanismo e operação historiográfica no IHGB republicano (1889-1933)”. (Tesis, IFCH,UFRGS, 2019). Ana Paula Barcelos, “Diálogos sobre a escrita da história. Ibero-americanismo, catolicismo (des)qualificação e alteridade no Brasil e na Argentina” (Tesis, UFF, 2011). Lucia Maria Paschoal Guimarães, *Da escola palatina ao silogeu: Instituto Histórico e Geográfico Brasileiro* (Rio de Janeiro: Museu Da República, 2007).
37. La Junta de Historia y Numismática Americana pasó a llamarse Academia Nacional de Historia de Argentina en 1938, por decreto del presidente Agustín Justo.
38. La incidencia de los intelectuales argentinos es menor porque el *Boletim da Junta de História e Numismática Americana* solo logró tener una periodicidad recién en 1924 y menos constante que la *Revista do IHGB*.
39. Fernando Devoto y Nora Pagano, *Historia de la historiografía argentina* (Buenos Aires: Sudamericana, 2009).
40. Alexandra Pita, “Pan-americanismo y nación”, *Anuario IEHS* 32, n.º 1 (2017): 136.
41. Raquel Paz Santos, *Um novo olhar sobre o país vizinho* (Rio de Janeiro: Multifoco, 2012).
42. Amado Luis Cervo y Wolfgang Dopcke, *Relações internacionais dos países americanos: Vertentes da história* (Brasília: UNB, 1994).
43. Luis Carlos Villafañe Santos, *O Brasil entre a Europa e a América: o Império e o interamericanismo (do Congresso do Panamá à Conferência de Washington)* (São Paulo: UNESP, 2004).
44. Los códigos están comentados. En Python, todo lo que está en la misma línea después de # no es leído por el intérprete del programa.
45. Adaptado de Christoph Rackwitz. “Tesseract ocr array of images to text csv”. Stackoverflow en <https://stackoverflow.com/questions/69239883/tesseract-ocr-array-of-images-to-text-csv>
46. Algoritmo desarrollado por nosotros.