

#noviembreHD. Cuarto congreso de la Asociación Argentina de Humanidades Digitales (AAHD). Asociación Argentina de Humanidades Digitales, Buenos Aires, 2021.

Trabajando con Voyant Tools.

José Luis DE PIERO.

Cita:

José Luis DE PIERO (2021). *Trabajando con Voyant Tools.*
#noviembreHD. Cuarto congreso de la Asociación Argentina de Humanidades Digitales (AAHD). Asociación Argentina de Humanidades Digitales, Buenos Aires.

Dirección estable: <https://www.aacademica.org/noviembrehd/30>

ARK: <https://n2t.net/ark:/13683/ehed/gnH>



Esta obra está bajo una licencia de Creative Commons.
Para ver una copia de esta licencia, visite
<https://creativecommons.org/licenses/by-sa/4.0/deed.es>.

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: <https://www.aacademica.org>.

Trabajando con Voyant Tools

Coordinación

Gimena del
Rio Riande

Romina De
León

José Luis DE PIERO
jl.depiero@filo.unt.edu.ar
INVELEC – CONICET – UNT

Para los estudios en Humanidades el análisis textual es una técnica fundamental, pero no por ello sencilla. En la actualidad existen dos cuestiones básicas que un investigador debe preguntarse antes de abordar un corpus: qué datos puede obtener para analizar y, luego, cómo leer esos datos. Para estos interrogantes existen ya respuestas brindadas desde el campo de las Humanidades Digitales. En el primer caso, Schöch (2013) propone que, para colecciones de corpus diferentes, existen datos que deben volverse cada vez más grandes y cada vez más inteligentes (*smart big data*), esto es, datos en grandes cantidades y que además posean propiedades suficientemente relevantes para los abordajes humanísticos.

Para el segundo problema, Moretti (2013) propone lo que denomina lectura distante (*distant reading*). Se trata de un análisis de un gran conjunto de datos a través de distintas estrategias provenientes de áreas tales como la historia, la geografía y la teoría evolutiva aplicadas a estos textos, lo que devuelve como resultado nuevas visualizaciones de los objetos analizados distintas a la lectura lineal.

Con estas ideas en mente, como un brevísimo y acotado marco teórico, presentamos una herramienta que puede ayudarnos a estos análisis: a leer, obtener datos y visualizarlos de una manera diferente. Se trata de Voyant Tools¹, una herramienta elaborada por el recientemente fallecido Stéfan Sinclair y Geoffrey Rockwell.

Voyant Tools es, de acuerdo con la información que ofrece su sitio web, “un entorno de lectura y análisis basado en web para textos digitales”. Se presenta como un proyecto académico que busca facilitar las prácticas de lectura y de interpretación para estudiantes e investigadores de las Humanidades Digitales. En líneas generales, este entorno permite analizar textos encontrados en la web o editados por el usuario, añadir paneles interactivos que per-

miten mostrar las propias lecturas y análisis a los demás y desarrollar herramientas propias.

Una de las principales ventajas de Voyant Tools es su interfaz intuitiva y cuidadosamente sencilla. La pantalla principal muestra un espacio para copiar una url, pegar un texto plano o cargar un archivo de texto editado por el autor. Permite, desde el comienzo, determinar el idioma automática o manualmente entre trece disponibles (alemán, arábico, bosnio, checo, croata, español, francés, inglés, italiano, japonés, portugués y serbio). Además, ofrece una serie de opciones para el cargado de metadatos y para la personalización del análisis si es que fuera requerido.

Una vez cargado se ofrece un panel con un conjunto de cinco herramientas por defecto. Cada una permite visualizar distintos tipos de datos del texto. Este panel es integralmente modificable y permite incluir distintos espacios de trabajo de acuerdo con las necesidades específicas de cada investigación. Presentaremos a continuación las que se muestran por defecto y mencionaremos algunas de las que aparecen ocultas al final.

La primera de ellas es Cirrus, que genera automáticamente una nube de palabras y muestra en mayor tamaño aquellas que aparecen una mayor cantidad de veces en el texto. Se puede visualizar como nube gráfica y elegir cuántas palabras incluir en la nube (entre 25 y 500), se puede visualizar la tabla de conteo de términos y los enlaces que muestran las relaciones de las palabras en proximidad junto a los términos claves. Entre las opciones, quizás la principal es la que permite editar la lista de palabras excluidas que contempla vocablos comunes normalmente no cargados semánticamente (como pronombres, determinantes o conjunciones).

La siguiente función es la de Reader, que muestra el texto plano cargado y la posibilidad de mostrar los términos más frecuentes como racimo, donde el tamaño de cada círculo representa mayor frecuencia absoluta de aparición de cada término.

La frecuencia relativa puede observarse en un gráfico lineal en Trends. La línea indica cómo varía esta frecuencia a lo largo de un corpus o de un mismo documento.

La herramienta Summary brinda un conjunto de datos básicos del texto: total de palabras, cantidad de palabras únicas, densidad de vocabulario, promedio de palabras por oración y palabras más frecuentes del corpus. Esta herramienta es particularmente útil cuando se trabaja con un corpus de varios elementos, ya que permite comparar la información entre los distintos documentos.

Finalmente, la herramienta Contexts permite visualizar los términos anteriores y posteriores a un término seleccionado a través del corpus y visualizarlo como líneas de burbuja o en concurrencias correlacionadas.

Entre las herramientas que no se muestran por defecto, algunas que permiten visualizar textos de distintas maneras son Mandala, que traza relaciones entre los términos más relevantes; WordTree, que visualiza concurrencias de términos en contexto, respecto de un término elegido; y Knots, que permite visualizar los términos más frecuentes como líneas y dobleces, donde cada doblez representa una ocurrencia.

Entre las herramientas experimentales mencionaremos DreamScape, que posibilita visuali-

¹ Accesible desde: <http://voyant-tools.org>.

zar ubicaciones y recorridos en un mapa, y RezoViz, que permite ver las relaciones entre personas, lugares y organizaciones. Estas dos últimas son quizás las más limitadas de las herramientas. En primer lugar, porque los diccionarios con los que realizan las búsquedas están en inglés y además no están completamente optimizados: pueden confundir notablemente nombres comunes de personas y lugares, y brindar información errónea.

Como observamos, la gran cantidad de herramientas integradas, entre las que se incluyen varias más no mencionadas en esta reseña, y la posibilidad de personalizar los paneles y la gratuidad del acceso son, quizás, las principales ventajas que ofrece este entorno.

Entre las desventajas podríamos pensar en los tiempos de carga que se incrementarán a medida que el corpus de textos sea mayor y las visualizaciones sean más complejas, lo que provoca, en ocasiones, errores que vuelven necesario actualizar la página o contar con una mayor velocidad de conexión a internet.

Otra cuestión que consideramos importante mencionar es la necesidad de conocer adecuadamente el corpus con el que se está trabajando. Si bien la herramienta muestra una amplia cantidad de información y de datos a primera vista, es necesario tener presente que a veces, sin una adecuada edición de los términos excluidos, esta información puede resultar poco relevante o inadecuada para ciertos tipos de preguntas. Como sucede en cualquier investigación, es fundamental tener en claro, antes de comenzar a trabajar, qué tipos de datos se están necesitando obtener.

Pensamos, como ejemplo de esto, algunos análisis críticos de textos en los que es fundamental la inclusión de los pronombres personales y demostrativos, ya que sus apariciones y contextos pueden brindar mucha información sobre el posicionamiento discursivo del autor de un texto.

Considerando las salvedades antes mencionadas, para concluir podemos afirmar que Voyant Tools es una herramienta dinámica, accesible y que ofrece una gran variedad de formas de visualizar, entender y analizar un corpus determinado. Por último, el hecho de que su código sea abierto permite a quien lo desee modificar o instalar la herramienta, según las necesidades de cada proyecto. Sin dudas, será interesante observar la evolución de este entorno con el tiempo y según surjan nuevas necesidades y posibilidades para el abordaje de textos.

REFERENCIAS BIBLIOGRÁFICAS

Moretti, F. (2013). *Distant Reading*. Verso.

Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities*, 2(3). <http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/>