

# **Ciencia de datos para las Humanidades Digitales**

Exploración de un corpus de hechicería

Rodrigo S. Cortez Madrigal

Escuela Nacional de Estudios Superiores

Campus Morelia

20 de septiembre de 2020

# Breve introducción a las Humanidades Digitales

Las humanidades digitales consisten en repensar el quehacer humanístico desde el punto de vista de todo el conjunto de las tecnologías mas recientes. Se trata de una disciplina en donde convergen la computación y las humanidades y que es inherentemente Interdisciplinaria y colaborativa. Combina las metodologías propias de las humanidades con técnicas computacionales para producir nuevas aplicaciones y técnicas que permiten nuevos tipos de investigación y enseñanza. Así pues, se genera una relación en la que las humanidades hacen uso de la herramientas que la tecnología provee para la investigación y al mismo tiempo las tecnologías son sometidas a un cuestionamiento humanístico. Es decir en esta manera de simbiosis ambas disciplinas se benefician mutuamente y están en constante transformación, por lo que dar una definición exacta y permanente de lo que son las Humanidades Digitales es un tarea difícil. Con la migración de materiales a entornos digitales surgen nuevos retos relativos, pero también surgen nuevas áreas de oportunidad. Este nuevo enfoque cambia la cultura de la labor de las humanidades así como las preguntas que se pueden hacer a los materiales que conforman el objeto de estudio de la investigación humanística. [1] La investigación en humanidades digitales es reconocer y prever los grandes cambios que ya se dieron y todos aquellos que están por venir en mayor medida debido a los grandes avances de la tecnología y la generación de volúmenes de datos cada vez mas grandes.

Las ciencias y las humanidades siempre han sido hermanas en su afán por el rigor intelectual y la libre investigación. [1] A pesar de sus diferencias no están en disputa, pero aun dicho esto, las humanidades digitales tratan de redefinir los límites entre las humanidades y las ciencias. El desarrollo de esta área busca también expandir la audiencia e incrementar el impacto de la investigación, enseñanza y producción. Crear nuevas preguntas que lleven a nuevo conocimiento y retomar algunas otras que se hayan olvidado en el camino. [1] En ese sentido, las humanidades digitales no buscan reemplazar el quehacer humanístico tradicional, sino todo lo contrario.

La cuestión primordial entonces en el quehacer de las Humanidades Digitales es ¿Cómo retomar y dar nueva forma a los métodos tradicionales en términos de técnicas computacionales?

## Tipos de estudio

El trabajo en las humanidades digitales utilizan una gran colección de métodos y los podemos encontrar en distintos formatos. El Laboratorio de Humanidades Digitales de Yale propone cuatro métodos principales para la investigación. [2]

El primero que es el Análisis de Textos consiste en técnicas para estudiar patrones en grandes colecciones de escritos. La minería de textos nos permite modelar el lenguaje de tal manera en que podemos encontrar las firmas lingüísticas de autores o patrones generales en colecciones de textos para detectar temas, por poner algunos ejemplos. La tarea de Análisis Visual implica la utilización de algoritmos que permitan estudiar patrones visuales en colecciones de imágenes. Estos complejos algoritmos permiten calcular la semejanza entre imágenes, identificar objetos o incluso obtener el contexto de una escena utilizando inteligencia artificial. El análisis espacial permite investigar cómo es que algún objeto ha cambiado a través del espacio-tiempo. Por poner algunos ejemplos es posible estudiar cómo ha cambiado un lugar o cómo se ha desplazado geográficamente algún grupo de individuos a lo largo del tiempo.

Finalmente uno de los estudios más interesantes que proponen es el estudio de Redes. Una red para los matemáticos y computólogos puede ser representada como un gráfo. Se trata de una estructura de datos que representa relaciones entre objetos de estudio, donde estos pueden ser prácticamente cualquier cosa, libros, personas, lugares o palabras por poner algunos ejemplos. Existen algoritmos que trabajan sobre estas estructuras de datos que permiten encontrar patrones en estas relaciones.

El trabajo en el campo de las Humanidades Digitales es en mayor medida orientado a proyectos. Esto quiere decir que en la plantación del proyecto se proponen preguntas relacionadas a los datos que se quieren estudiar, se forma el corpus con los materiales o los datos en general, se utilizan herramientas o métodos computacionales, para analizar y visualizar y finalmente se utilizan herramientas de diseminación digital para compartir el trabajo. [2] Este ultimo paso quiere decir que no solamente se utiliza un producto publicable, sino una plataforma web o un repositorio digital para conectar con la audiencia.

## Hechizos

Nuestra idea original consistía en precisamente utilizar esta metodología y herramientas anteriormente mencionadas para explorar un corpus de hechicería contemporánea que fue recolectada por la Dra. Cecilia López Ridaura, profesora de tiempo completo de la Licenciatura en Literatura Intercultural y jefa del departamento editorial de la ENES Morelia. Nos interesaba principalmente así crear herramientas para la visualización de estos datos y explorar la relación la composición química y textual de los hechizos obtenidos.

### Descripción de los datos

El corpus está conformado por 154 hechizos recolectados en distintos mercados de la Ciudad de México y Morelia. Un hechizo esta descrito de la siguiente manera.

Para obtener la composición química de los hechizos acudimos al Laboratorio de Microscopía de la ENES Morelia donde utilizaron Difractometría de Rayos X y un Microscopio Electrónico de Barrido (MEB) par obtener la información anteriormente mencionada. Técnicas que van más allá del ámbito de esta ponencia y se abordarán en otra ocasión.

<b>Hechizo</b>
Nombre
Oración
Objeto
Procedencia
Medidas
Datos Adicionales
Voz
Elementos químicos

## Visualización

Una de las tareas claves para las Humanidades Digitales que se mencionaba anteriormente era la de expandir la audiencia de la investigación, para lograr eso uno de los recursos más importante y de mayor impacto es el de la visualización de datos. Esto permite hacer aun más accesible los contenidos del proyecto en Humanidades Digitales para todo aquel interesado sin ser necesariamente experto en el tema, se trata por lo tanto también de una labor de divulgación. Es tanto como un arte como una ciencia. [3] En nuestro caso se optó en primer lugar por una plataforma en la Web que permitiese explorar el contenido de los hechizos a cualquiera que tuviese interés en el contenido de una manera fácil y accesible. La página web también tiene la finalidad de alojar los estudios posteriores que aún tenemos por hacer.

En la página principal podemos encontrar un cuadro en el que los hechizos se representan en una nube de puntos en la que cada color representa la pertenencia del hechizo a una categoría seleccionable como la procedencia, por ejemplo. En la página de inicio también se encuentra acceso a información adicional sobre el proyecto, los reportes de estudio, el repositorio donde se aloja el código y a los datos de contacto de los responsables del proyecto.

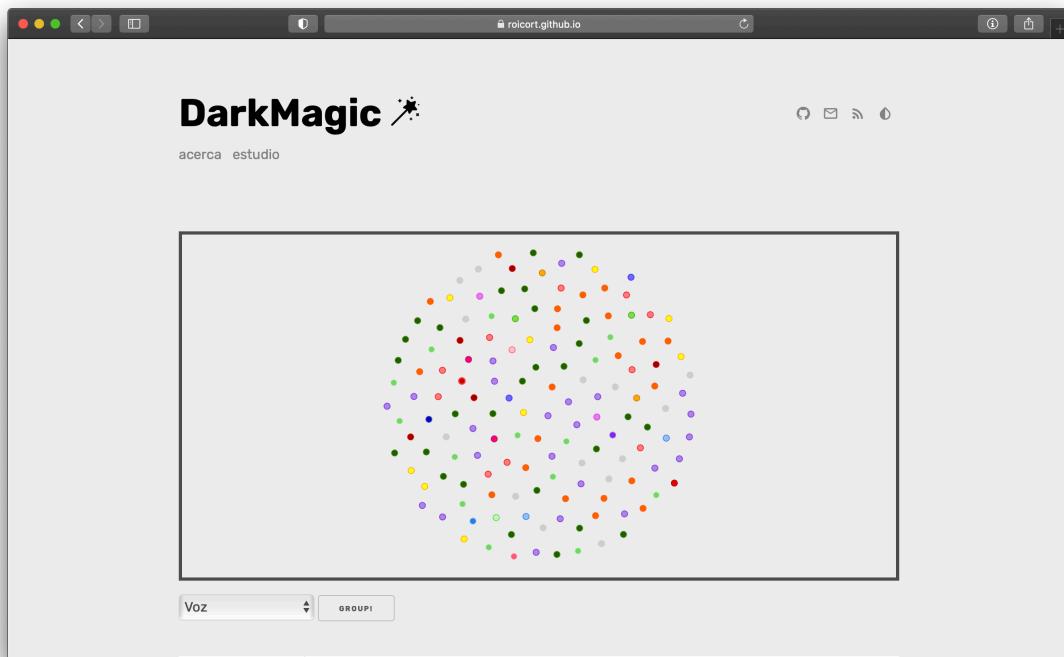


Fig. 1

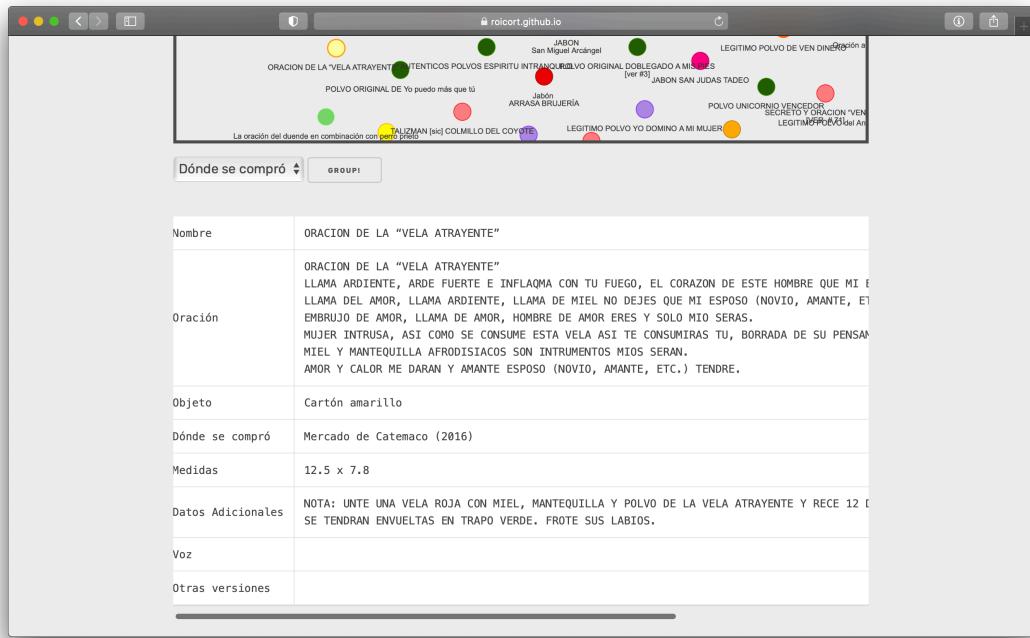


Fig. 2

<sup>1</sup> Página de inicio de la web de proyecto - Disponible temporalmente en: <https://roicort.github.io/DarkMagic>  
Nombre clave: DarkMagic - No representa el nombre final de la plataforma web

<sup>2</sup> Visualización de datos en la pagina web del proyecto. En esta imagen podemos observar la información desplegada al seleccionar un hechizo.

A continuación algunos ejemplos de los estudios que he realizado hasta el momento.

### Análisis Químico

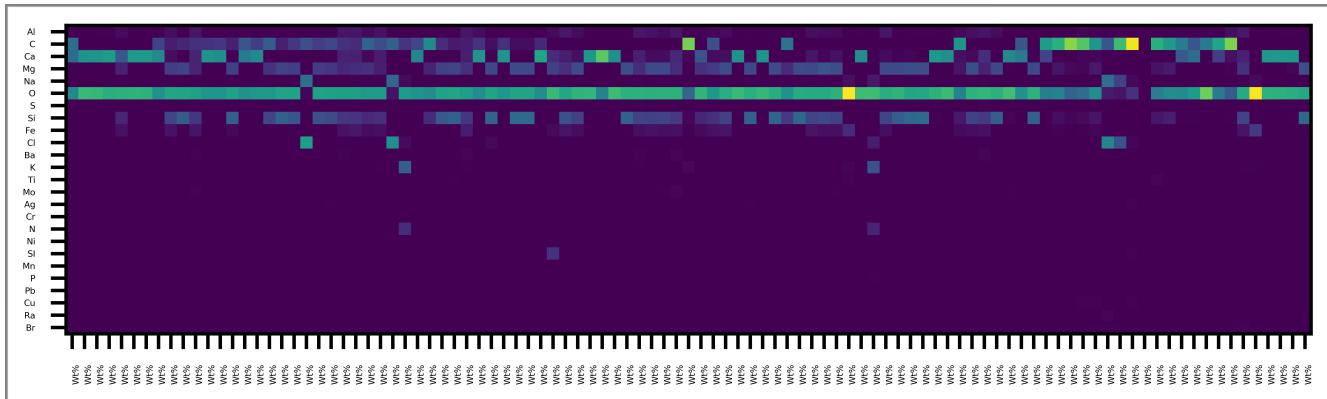


Fig. 3

En la Fig. 3 es posible observar el conjunto de elementos químicos que componen al corpus de los hechizos. En la gráfica cada columna representa un hechizo y cada fila representa el porcentaje de composición de cada elemento, donde el espectro de color morado amarillo representa el rango de 0 a 100%

### Lenguaje



Fig. 4

<sup>3</sup> Representación gráfica de los elementos encontrados y el porcentaje de composición química por hechizo

En la Fig.4 podemos observar una nube de palabras de los verbos más frecuentes utilizados en las oraciones que conforman los hechizos. Donde el tamaño de las palabras representa la frecuencia de aparición en el corpus. Utilizando Inteligencia Artificial podemos crear modelos de lenguaje que nos permiten diferenciar entre las entidades en el texto.

## Conclusión

Es importante mencionar que este trabajo aún están en desarrollo y queda mucho trabajo por hacer. Esperamos mejorar la visualización y realizar más experimentos con los datos que nos permitan estudiar la relación entre la composición química y textual de los hechizos. Creemos que la plataforma que estamos creando permitirá la difusión, estudio e interés en este tipo de datos. También buscamos en este espacio escuchar retroalimentación de los expertos en otras áreas del conocimiento que nos permitan enriquecer el proyecto y sobre todo seguir trabajando desde la interdisciplinar e invitar a otros a hacer lo mismo.

## Agradecimientos

Responsables del Proyecto PAPIIT IA106620

- Dra. Marisol Flores Garrido
- Dr. Luis Miguel García Velásquez

Otros estudiantes en el proyecto

- Hector De la Rosa
- Nohely Fierros
- Carlos Cortés

Escuela Nacional de Estudios Superiores UNAM Campus Morelia

Licenciatura en Tecnologías para la Información en Ciencias

## Bibliografía

- [1] A. Burdick, Ed., *Digital Humanities*. Cambridge, MA: MIT Press, 2012.
- [2] Yale University Library, «Digital Humanities Lab», *Digital Humanities Lab*. <https://dhlab.yale.edu/guides.html>.
- [3] M. Aparicio y C. J. Costa, «Data visualization», *Commun. Des. Q. Rev*, vol. 3, n.º 1, pp. 7-11, ene. 2015, doi: [10.1145/2721882.2721883](https://doi.org/10.1145/2721882.2721883).