

Título del proyecto:

Procesar y comprender información: generación de conocimiento para las Humanidades desde las Humanidades Digitales

1. Problema de investigación

Una de las grandes desventajas que enfrentan los investigadores y estudiantes en el campo de las Humanidades tiene que ver con la aparente falta de experimentación y aporte de datos duros que evidencien, entre otras cosas, la generalización de resultados o la aplicación de los hallazgos más allá del contexto propio de la investigación en la que se obtienen. Ello conlleva, en algunos casos, a generar una visión en la que el papel de las Humanidades en la construcción de conocimiento sea considerado poco relevante. Considérese, por ejemplo, la discusión sobre la pertinencia de incluir las Humanidades como parte constitutiva del quehacer científico en el CONAHCYT.

El problema que esto supone no está relacionado exclusivamente con los fenómenos de estudio que interesan a las Humanidades ni con los marcos conceptuales y metodológicos que las determinan (i.e. contenido disciplinar), sino, desde nuestra perspectiva, con el desconocimiento de modelos, recursos y herramientas (i.e. contenido multi y/o transdisciplinar) que permitirían a los investigadores de estas áreas expandir los alcances de sus estudios, tanto en términos de la obtención de datos, como en términos de la capacidad de experimentación, representación de características y aplicación de métricas para dar mayor sustento a sus resultados y, en lo posible, generalizarlos.

En este contexto, quienes ven mayormente limitados sus intereses de investigación son los estudiantes, puesto que están en un proceso de adquisición de conocimientos enmarcado en programas de estudio que tienden a focalizar contenidos disciplinares, pero que pocas veces consideran contenidos de carácter multi o transdisciplinar que les permitan adquirir competencias y habilidades que tradicionalmente se ven como externas a sus disciplinas.

Con base en este problema, en este proyecto se plantea las siguientes preguntas de investigación.

¿Es factible desarrollar en el estudiante de los programas de Humanidades un método de trabajo transdisciplinar, a través de las Humanidades Digitales (HD), que le permita perfeccionar y dar mayor alcance a sus investigaciones?

¿Qué tipo de necesidades deberían cubrir las HD en función de los variados fenómenos de estudio que interesan a las Humanidades?

¿Cuáles serían las evidencias formales que permitan ponderar las ventajas y/o desventajas de las HD en las investigaciones de los estudiantes de Humanidades?

2. Antecedentes y Estado del arte

En una visión muy general, las HD pueden considerarse como un campo en el que se busca “entender el impacto y la relación de las tecnologías de cómputo en el quehacer de los investigadores de las Humanidades” (Galina Russell, 2011). En esta interrelación, las HD no pretenden desplazar o sustituir la mirada experta del humanista, sino complementarla, proporcionando nuevas herramientas para comprender los fenómenos que la circundan, obteniendo más elementos y perspectiva para el análisis (Gabi, 2022). A partir de lo anterior, (Burdick, 2012) destaca que las actividades fundamentales de las HD refieren a la curación, el análisis, la edición y el modelado, principalmente, del lenguaje. Asimismo, señala que para poder llevar a cabo estas actividades es necesario el procesamiento de datos, la obtención y generación de metadatos, la descripción, organización, clasificación y visualización de datos, entre otros.

Si bien, la historia de las HD es reciente (algunos autores consideran el Índice Thomasticus de 1974 como el origen de estas), su aplicación en diversos proyectos ha evidenciado el alcance que pueden lograr en el campo de las Humanidades. Véanse, por ejemplo, el trabajo de (Warwick, 2012b) para describir las características de los usuarios de las HD; el de (Terras, 2012) y de (Ross, 2012) en los que se detalla el diseño para la creación de recursos digitales y el uso de redes sociales en HD, respectivamente; o el de (Warwick, 2012a) que trata sobre los esfuerzos que han hecho instituciones como el King’s College London, organismos internacionales, tales como la Modern Language Association (<https://www.mla.org/>) o Centernet (<https://dhcenternet.org/>), en relación con la implementación de las HD para profundizar en problemáticas que antes eran muy difíciles de tratar o que implicaban un esfuerzo humano colosal.

En el contexto nacional, a pesar de que un estudio realizado por el Center for Digital Humanities de la University College London para Cuantificar las Humanidades Digitales (Terras, 2011) mostraba que en México no había ni individuos ni recursos, en el documento de (Galina Russell, 2011) se reportan algunos proyectos que, si bien no estaban catalogados en un inicio como de HD, sí han generado resultados que han impactado en el área y que han permitido que esta se constituya formalmente. Asimismo, existe la Red de Humanidades Digitales (<http://humanidadesdigitales.net>) que se ha encargado de encausar, difundir e, incluso, normar el trabajo que diferentes investigadores, instituciones y organismos están realizando en el país.

Con base en lo anteriormente expuesto, este proyecto se justifica a partir de las problemáticas y desafíos que supone a la sociedad la enorme cantidad de información que existe hoy en día, así como su correcta interpretación para, por ejemplo, la mejor toma de decisiones. En este sentido, las palabras de Burdick et al. sintetizan de forma clara tales problemáticas, además de que resaltan el rol que tendrá el humanista que puede aplicar HD a su labor:

The social, political, and ecological challenges of the 21st century demand significantly more than textual analysis or recitations of inherited content. These problems (and opportunities) will need people trained to create synthetic responses, rich with meaning and purpose, and capable of communicating in a range of appropriate media, including but not limited to print (Burdick, 2012).

3. Fundamentación teórica

El proyecto, en su vertiente más general, retoma los fundamentos de las HD, pero en la parte más específica, adopta los fundamentos conceptuales y metodológicos de áreas especializadas, tales como la Lingüística de Corpus y el Procesamiento del Lenguaje Natural.

La lingüística de corpus es una disciplina que se caracteriza por realizar análisis de contenido textual mediante el uso de software especializado. La metodología de esta disciplina se orienta hacia el análisis empírico del lenguaje (O'Keeffe & McCarthy, 2010) a través de la descripción, cuantitativa y cualitativa, de las formas naturales en que se verbaliza una lengua. Para lograrlo, se hace uso de técnicas que permiten representar información variada y a diferente nivel, por ejemplo, información frecuencial, palabras clave, coligaciones, colocaciones, concordancias y demás. Asimismo, los programas de cómputo que agrupan estas técnicas implementan diversos algoritmos para reconocer patrones y comportamientos que pueden describirse, explicarse y argumentarse por medio de un conjunto muy amplio de métricas estadísticas y de probabilidad, tales como χ^2 , perplejidad, entropía, función de verosimilitud, entre otras (Oakes, 1998; Manning & Schütze, 1999).

Por otra parte, el Procesamiento del Lenguaje Natural o PLN tiene como propósito modelar cualquier tipo de interacción humano-computadora a través del lenguaje natural (Bird et al., 2009), sin importar si este es escrito o hablado. De acuerdo con varios expertos, el PLN se constituye a partir de la interacción entre disciplinas e implica diferentes campos, por ejemplo, la lingüística, las ciencias de la computación o la psicología (Jurafsky & Martin, 2007). En términos formales, el objetivo del PLN es reproducir la transmisión de información de la forma más natural posible entre un humano y una computadora a partir de modelar la producción de los hablantes y la interpretación de los oyentes (Hausser, 1998). Este objetivo se ha logrado, en parte, debido al desarrollo de algoritmos que pueden recuperar el significado de una expresión a partir del contenido lingüístico de esta (Clark et al., 2010). Cabe destacar que, dada su naturaleza transdisciplinaria, el PLN adopta técnicas, enfoques, herramientas y aproximaciones que provienen de áreas como la inteligencia artificial, las ciencias de la computación, la estadística, la lingüística teórica, entre otras.

4. Hipótesis

El uso consciente de las HD en las investigaciones de los estudiantes de los programas de Humanidades que se imparten en la Facultad de Lenguas y Letras de la UAQ, supone un trabajo metodológico y descriptivo más profundo para el análisis de sus objetos de estudio y la generalización de sus hallazgos.

5. Objetivos

El objetivo del proyecto es evaluar la eficacia de las HD para generar conocimiento en el campo de las Humanidades, en específico, en los programas de licenciatura y posgrado que se imparten en la Facultad de Lenguas y Letras de la UAQ.

Para lograr este objetivo, se consideran los siguientes objetivos específicos:

- a) Identificar qué tipo de problemáticas tienen los estudiantes de los programas en Humanidades que se imparten en la Facultad de Lenguas y Letras de la UAQ en cuanto al desarrollo de sus investigaciones, particularmente, en relación con la obtención, procesamiento y análisis de datos.
- b) Categorizar las problemáticas en función de los alcances que estos programas educativos tienen.
- c) Establecer una correspondencia entre la forma o formas en que los estudiantes tradicionalmente han resuelto estas problemáticas y las cuatro actividades centrales de las HD: curación, análisis, edición y modelado.
- d) Obtener, mediante enfoques tradicionales de Humanidades, datos relativos a tiempo empleado y precisión de resultados en tareas de extracción de información, variabilidad y riqueza de análisis, así como de presentación y alcance de hallazgos.
- e) Realizar las mismas tareas del objetivo previo con un enfoque de HD.
- f) Comparar los alcances de cada enfoque y determinar las ventajas y desventajas de cada uno.
- g) Sistematizar los resultados para desarrollar un taller que ejemplifique, a partir de investigaciones puntuales, los alcances que conlleva la aplicación de las HD en proyectos de traducción, de análisis literario, de análisis del discurso, así como de generación de recursos en lenguas originarias mexicanas. Todos ellos, parte fundamental del trabajo humanístico que se desarrolla en la FLL.

6. Metodología

El proyecto propuesto es de enfoque mixto con alcance descriptivo y correlacional. Los datos se obtendrán a partir de las investigaciones que realicen los estudiantes que participen en las diferentes pruebas del proyecto. La información personal y de diseño de investigación que se recabe en esta etapa solo se usará con fines de análisis de acuerdo con el objetivo del proyecto. Todos los estudiantes que

participen deberán firmar un documento de consentimiento en el que se especifica el objetivo del proyecto, el uso de los datos que aportarán, la protección de sus datos personales conforme a la legislación mexicana, así como la revocación de su consentimiento. Este consentimiento se incluye en el presente documento como Anexo 1. El resguardo de esta información será en archivo electrónico a cargo del responsable del proyecto.

Adicionalmente, para ejemplificar el trabajo con HD y poder comparar los resultados que se describen en el objetivo específico 6, se crearán corpus ad hoc en el ámbito de la traducción, el análisis literario y el análisis del discurso. Para la construcción de estos corpus se respetarán los derechos de las plataformas de las cuales se recupere la información, tales como políticas de privacidad y uso de datos, así como de descarga a través de APIs oficiales (si aplica), las cuales verifican que se cumple con los criterios necesarios para tener acceso a los contenidos publicados por sus usuarios. Asimismo, para garantizar la confidencialidad de los datos una vez que se generen los corpus, estos solo podrán ser compartidos con fines académicos y de investigación, respetando en todo momento las políticas de cada plataforma.

Para el caso del taller que se compromete como entregable, solo se hará el registro de los asistentes (nombre, programa, correo electrónico) con fines estadísticos. Ningún otro dato será recabado. El resguardo de esta información será en archivo electrónico a cargo del responsable del proyecto.

Los recursos materiales que se usarán en el proyecto se obtendrán del financiamiento solicitado. En cuanto a los recursos materiales que se solicita a los participantes, será un equipo de cómputo. Si el alumno no tiene uno disponible para llevarlo a la facultad, tanto para las pruebas como para el taller, se usará el laboratorio de cómputo de la FLL.

El análisis de datos se llevará a cabo mediante el uso de software libre para minimizar costos de licencias propietarias, así como para garantizar, en lo posible, el uso multiplataforma, es decir, que el software se pueda ejecutar en varios sistemas operativos.

7. Resultado esperados y aplicaciones

Los resultados que se esperan alcanzar con este proyecto se sintetizan a continuación:

- i. Identificación de problemáticas relacionadas con la obtención, procesamiento y análisis de datos que deben resolver los alumnos de los programas ofertados en la FLL.
- ii. Aplicación concreta de técnicas, herramientas y recursos de las HD para resolver las problemáticas antes mencionadas.

- iii. Reducción de los tiempos para realizar algunas tareas de obtención, procesamiento y análisis de datos que son comunes a las investigaciones que desarrollan los estudiantes de la FLL.
- iv. Expansión del alcance exploratorio, descriptivo y correlacional de las investigaciones de los estudiantes de la FLL mediante la aplicación consciente de las HD.
- v. Diversificación de posibilidades de aplicación de las HD, tanto en el plano laboral como en de formación a nivel posgrado.
- vi. Asimismo, entrega de los productos comprometidos: una tesis de maestría, un capítulo de libro, una presentación en congreso y un taller para los alumnos de la FLL y estudiantes interesados de otras facultades.

Referencias

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly.

Burdick, A. (2012). *Digital humanities* / Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, Jeffrey Schnapp. 141.

Clark, A., Fox, C., & Lappin, S. (2010). *The Handbook of Computational Linguistics and Natural Language Processing*. En *Blackwell Handbooks in Linguistics*. John Wiley & Sons.

Gabi, K. (2022). *Computation as Context: New Approaches to the Close/Distant Reading Debate*. *College Literature*, 49(1), 1–25. <https://doi.org/10.1353/LIT.2022.0000>

Galina Russell, I. (2011). ¿Qué son las Humanidades Digitales? *Revista Digital Universitaria*, 12(7), 3–11.

Hausser, R. (1998). *Foundations of Computational Linguistics*. Springer-Verlag.

Jurafsky, D., & Martin, J. (2007). *Speech and Language Processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice Hall.

Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.

Oakes, M. (1998). *Statistics for Corpus Linguistics*. Edinburgh University Press.

O'Keeffe, A., & McCarthy, M. (Eds.). (2010). *The Routledge handbook of corpus linguistics* (1a ed.). Routledge. <https://www.routledge.com/The-Routledge-Handbook-of-Corpus-Linguistics/OKeeffe-McCarthy/p/book/9780367076382>

Ross, C. (2012). Social media for digital humanities and community engagement. En C. Warwick, M. Terras, & J. Nyhan (Eds.), *Digital Humanities in Practice* (pp. 23–46). Facet. <https://doi.org/DOI: 10.29085/9781856049054.003>

Terras, M. (2011). Infographic: Quantifying Digital Humanities | UCL UCLDH Blog. <https://blogs.ucl.ac.uk/dh/2012/01/20/infographic-quantifying-digital-humanities/>

Terras, M. (2012). Digitization and digital resources in the humanities. En C. Warwick, M. Terras, & J. Nyhan (Eds.), *Digital Humanities in Practice* (pp. 47–70). Facet. <https://doi.org/DOI: 10.29085/9781856049054.004>

Warwick, C. (2012a). Institutional models for digital humanities. En C. Warwick, M. Terras, & J. Nyhan (Eds.), *Digital Humanities in Practice* (pp. 193–216). Facet. <https://doi.org/DOI: 10.29085/9781856049054.010>

Warwick, C. (2012b). Studying users in digital humanities. En C. Warwick, M. Terras, & J. Nyhan (Eds.), *Digital Humanities in Practice* (pp. 1–22). Facet. <https://doi.org/DOI: 10.29085/9781856049054.002>