BOWEN SAVANT, Sarah, ROMANOV, Maxim y THOMAS MILLER, Mathew.
Open Islamicate Texts Initiative (OpenITI), 2016.

<https://iti-corpus.github.io/>

Dirección

Clara Martínez
Cantón
Gimena del Río
Riande
Ernesto Priani

Secretaría

Romina De León

Reseña realizada por:

Abdenour PADILLO SAOUD

Consejo Superior de Investigaciones Científicas (CSIC)

padillosaoud@eeae.csic.es

Si bien la presencia de lo digital en ámbitos como las distintas filologías europeas cuenta con una trayectoria relativamente amplia, en el caso de los estudios árabes se presenta como un fenómeno incipiente, debido a la incorporación tardía y a la representación digital de la tipografía árabe¹ o la edición de los textos clásicos a partir de los manuscritos conservados plantean una serie de desafíos para la incorporación de los estudios islámicos al pujante campo de las Humanidades Digitales (HD).

Es en este contexto donde surge el proyecto Open Islamicate Texts Initiative (OpenITI). El objetivo principal del proyecto es crear el primer corpus digital de textos árabes y persas de época premoderna. Este objetivo, cuanto menos ambicioso, pretende establecer la infraestructura tecnológica necesaria para incorporar las nuevas formas de análisis y tratamiento computacional de la información a la ingente producción escrita de este periodo. Las posibilidades que abriría a los investigadores la consecución de este objetivo supondrían una revolución en nuestro campo de estudio, pero ¿en qué consiste exactamente este proyecto?

OpenITI parte de la colaboración conjunta de numerosos investigadores de diferentes instituciones. La dirección del proyecto está a cargo de Sarah Bowen Savant, profesora asociada de la Aga Khan University de Londres, Maxim Romanov, investigador de la Universidad de Viena y Matthew Thomas Miller, investigador de la Universidad de Maryland.

¹ Sobre esta problemática puede consultarse Izedin Mohamed Ali (2018).

En parte, en este proyecto confluyen los esfuerzos individuales que con anterioridad estos investigadores venían realizando en diferentes proyectos. La suma de estos esfuerzos amplía notablemente el horizonte del OpenITI, no solo en su aspecto cuantitativo, sino también en su capacidad de impacto por la implicación de diferentes instituciones internacionales.

El proyecto surge en el verano de 2016 con el objetivo inicial de albergar alrededor de 10.000 textos islámicos premodernos. Tras el último balance realizado por parte de los responsables del proyecto en 2018 se afirma que, hasta la fecha, se han incluido 4.200 textos (Miller, Romanov y Savant, 2018). Estos textos corresponden a obras de los principales géneros que florecieron en la cultura islámica desde el siglo IX hasta el XVI. En menor medida, el corpus actualmente también contiene obras de los periodos anteriores. OpenITI está alojado en el repositorio GitHub lo que permite un trabajo colectivo efectivo y seguro, pero a su vez plantea una serie de dificultades para los usuarios que no están familiarizados con el uso de esta herramienta. La mayoría de los textos han sido extraídos de repositorios online que han aparecido en las últimas décadas principalmente en el mundo islámico². Estos repositorios surgen con afán divulgativo, su objetivo es acercar de una manera sencilla y manejable la producción escrita del mundo islámico clásico a un público amplio y dispar. La temática de los textos que se incluyen en estos repositorios está relacionada con las ciencias islámicas. Otras tipologías de textos no incluidos en estas ciencias no son incorporadas a estos repositorios. Para solventar este sesgo, el equipo del proyecto desarrolló un sistema OCR que permite escanear de manera efectiva textos escritos en árabe (Kiessling, Miller, Romanov y Savant, 2017) y que será de utilidad para ampliar el corpus con textos no disponibles en estos repositorios.

Para el etiquetado del texto se ha desarrollado una versión compatible con la lengua árabe del sistema Markdown. Para ello se utiliza el editor de texto EditPad Pro, pues resulta especialmente útil por la posibilidad que ofrece de trabajo con lenguas que van de derecha a izquierda, como es el caso del árabe. Este sistema ha sido rebautizado como OpenITI mARKdown por los miembros del proyecto. Entendemos que esta elección, de momento, responde más a cuestiones prácticas relacionadas con la menor complejidad de este sistema y su idoneidad para lenguas como el árabe. Los miembros del proyecto tienen previsto, una vez finalizado el proceso de etiquetado, la conversión de estos textos al sistema TEI-XML.

Los textos están clasificados según el autor y la fecha en carpetas que hacen fácil el acceso a los mismos. Este corpus utiliza el protocolo Canonical Text Services (CTS)³ que resulta muy efectivo para la identificación y recuperación de textos o fragmentos de texto. Este protocolo asigna un URN (un nombre de recurso uniforme) y también un URI (identificador de recurso único), que proporciona las referencias canónicas permanentes a los textos o pasajes de

² Algunos de estos repositorios son: *Al-Maktaba Al-Shamela* (<http://shamela.ws/index.php>); *Waqfeya*: (<http://waqfeya.com/>) y *Shia Online Library* (<http://shiaonlinelibrary.com/>).

³ Una breve guía del CTS está disponible en: (<https://bit.ly/2YumdJK>).

texto y son utilizados por los Servicios de texto canónico (CTS) para identificar o recuperar pasajes de texto⁴.

Uno de los primeros problemas que encuentra el usuario que se acerca a este proyecto está relacionado con la información que aparece en las distintas webs vinculadas al mismo. La página principal ofrece información general y básica sobre los objetivos, implementación del proyecto y los participantes, si bien es cierto que esta contiene enlaces donde se puede ampliar la información, no encontramos vínculos directos con el repositorio donde se aloja el corpus. La web incluye la posibilidad de visualizar la información en árabe y persa, aspecto muy positivo y cuyo objetivo es llegar a los investigadores de estudios islámicos más allá de las fronteras de la academia occidental, pero estas opciones, pese a aparecer en la página de inicio, aún no están operativas. Por otra parte, existen otras herramientas vinculadas al proyecto, como acabamos de referir, que ofrecen una información más detallada, pero que no tienen un enlace directo a la web principal. Entre éstas destaca principalmente al-Raqmiyyāt. Digital Islamic History⁵ gestionada por uno de los encargados de OpenITI y uno de sus miembros más activos, Maxim Romanov. La página tiene una interfaz sencilla y manejable. Contiene información precisa y pertinente sobre los diferentes proyectos que surgen relacionados o a partir de OpenITI y lo que nos resulta más relevante, una descripción detallada de las distintas fases del proyecto y el funcionamiento del mismo, todo ello a través de enlaces que complementan y amplían la información. Ahora bien, el lenguaje utilizado en esta web no está pensado para aquellos usuarios no familiarizados con los conceptos computacionales.

En teoría, los miembros del proyecto apuestan de manera firme por el Open Access, sin embargo, en la web principal no aparece citada ninguna licencia de manera explícita. Este posicionamiento aparece en un manifiesto conjunto que pretende sentar los principios sobre los que se articula el proyecto y que paradójicamente no está publicado en acceso abierto, aunque en uno de los repositorios del proyecto aparece un enlace que remite al mismo de manera directa (Miller et al., 2018). En este artículo se afirma que, aparte de los beneficios que reporta a la comunidad investigadora el libre acceso a los recursos del corpus y a la producción que se genere a partir del mismo, los integrantes del equipo pretenden ampliar el alcance de este proyecto y de otros vinculados y derivados de OpenITI, con especial atención a la comunidad de investigadores y académicos de los países de mayoría islámica. Actualmente el contenido principal del corpus consiste en numerosos textos a los que se puede acceder de manera directa. Hasta donde hemos podido saber, se sigue trabajando en el procesamiento de textos con el fin de crear un software analítico específico y otra serie de herramientas que aún no han visto la luz y que, también en teoría, deberían aparecer en Open Access. A pesar de que esta omisión

⁴ Este proceso y el anterior están detallado con precisión en la web de uno de los miembros del proyecto, véase: <https://alraqmiyyat.github.io/OpenITI/>.

⁵ Accesible desde: <https://alraqmiyyat.github.io/>.

puede parecer una cuestión baladí para muchos, la mención explícita de esta condición en proyectos, textos y otra producción científica, resulta una herramienta útil para visibilizar y avanzar en la democratización del conocimiento, cuestión fundamental en las Humanidades Digitales en particular, y en el ámbito académico en general.

Pese a estas pequeñas deficiencias, que entendemos se irán subsanando a medida que el proyecto avance, OpenITI viene a dar respuesta a una necesidad cada vez más emergente en el campo de los estudios islámicos. Los objetivos que aquí hemos expuesto auguran una auténtica revolución en esta disciplina que exige la reflexión e implicación por parte de toda la comunidad investigadora. Este proyecto establece la infraestructura necesaria para gestionar las enormes posibilidades que nos ofrece la extensa producción escrita islámica del periodo premoderno, pero exige repensar las formas tradicionales de acercarnos a los datos, aún imperantes en nuestro campo de estudio, y que algunos teóricos han denominado lectura cercana (*close reading*). En este sentido, los arabistas asociados a las Humanidades Digitales, entre los que se encuentran muchos de los participantes del proyecto, han asumido los retos que implica estas nuevas formas de gestión de la información y, partiendo de la idea que propugnan algunos académicos como Moretti (2007) de la lectura distante (*distant reading*), están comenzando a cuestionar la validez de la lectura filológica tradicional para los datos que nos ofrecen los grandes corpus textuales como OpenITI⁶. Por otra parte, la repercusión de este tipo de proyectos en el arabismo español sigue estando muy limitada, pero comienzan a darse los primeros pasos⁷ hacia un proceso que parece imparable y que hace tiempo se instaló en nuestras sociedades.

Nuevos proyectos en HD están surgiendo junto a OpenITI que hacen presagiar un futuro prometedor para los estudios islámicos. Entre estos destaca el proyecto KITAB⁸ que proporcionar diferentes formas de análisis del corpus de OpenITI y que supone el siguiente estadio en las distintas fases del proyecto. Estas y otras iniciativas están ampliando la visión tradicional de los investigadores en este campo, reformulando las preguntas que veníamos haciéndonos y ofreciendo nuevas respuestas que convierten esta etapa en un desafío apasionante.

⁶ Significativo fue el congreso celebrado en la Universidad de Brown en 2015, titulado *Distant Reading and the Islamic Archive*, el programa se encuentra accesible desde: <https://bit.ly/2KYUmxU>.

⁷ El CSIC está siendo pionero en este aspecto. A finales de 2018 se celebró junto a otras instituciones el congreso “Magreb/al-Andalus hd”. Actualmente el proyecto de investigación que dirigen Maribel Fierro y Mayte Penelas está apostando por las HD a través de la formación de sus investigadores y la incorporación de investigadores relacionados con esta disciplina como el mencionado Maxim Romanov. Los datos del proyecto se encuentran accesible desde: <https://bit.ly/2UjgCaa>.

⁸ Accesible desde: <http://kitab-project.org/>.

REFERENCIAS BIBLIOGRÁFICAS

- Izedin Mohamed Ali, H. I. (2018). Arabic Typography Development and Technological Compatibility. *American Journal of Linguistics*, 6(2), 19-26. Recuperado de <https://bit.ly/2GuCyXJ> el 04/07/2019.
- Kiessling, B., Miller, M., Romanov, M. y Savant, S. (2017). Important New Developments in Arabographic Optical Character Recognition (OCR). *Al-'Usur al-Wusta*, 25, 1-13. Recuperado de <https://bit.ly/2YyL2Ei> el 04/07/2019.
- Miller, M., Romanov, M. y Savant, S. (2018). Digitizing the Textual Heritage of the Premodern Islamicate World: Principles and Plans. *International Journal of Middle East Studies*, 50(1), 103-109. doi:[10.1017/S0020743817000964](https://doi.org/10.1017/S0020743817000964).
- Moretti, F. (2007). *La literatura vista desde lejos*. Barcelona: Marbot ediciones.