

BOWEN SAVANT, Sarah et al. *KITAB Project. Knowledge, Information, Technology, and the Arabic Book.*

<http://kitab-project.org/>

**Dirección**  
Clara Martínez  
Cantón  
Gimena del Río  
Riande  
Ernesto Priani

**Secretaría**  
Romina De León

Reseña realizada por:  
Iria SANTÁS DE ARCOS  
Universidad Eclesiástica San Dámaso  
[iriasantas@hotmail.com](mailto:iriasantas@hotmail.com)

El Proyecto KITAB, dirigido por Sarah Bowen Savant en el centro Aga Khan University-ISMC de Londres es, en principio –aunque con ciertas reservas–, un proyecto ilusionante y esperanzador para los investigadores en estudios árabes e islámicos gracias a la multitud de posibilidades y campos de investigación que les ofrece. Veamos, a continuación, el gran objetivo que persigue esta iniciativa.

El proyecto comenzó en el año 2015 con el propósito de convertirse en una herramienta digital y en un foro abierto con el fin de discutir acerca de textos árabes y conocer en qué medida y de qué manera estos han sido utilizados de nuevo por otros autores en el periodo de tiempo comprendido entre los años 700-1500 a lo largo de todo el mundo islámico medieval. Esta cuestión es de gran relevancia, pues la tradición literaria medieval árabe, desde sus orígenes, se ha gestado con base en este principio que podríamos llamar comúnmente de *corta y pega*, en la medida en que un autor recurre a otros autores que le han precedido para narrar los datos del pasado de los cuales no ha sido testigo. Esta circunstancia, totalmente ajena, en aquella época, a lo que en la actualidad se considera plagio, ha sido fundamental en el desarrollo de las obras islámicas y árabes, ya que a través de esta técnica han llegado hasta nosotros textos completos o fragmentos de obras hoy perdidas, que, afortunadamente, se han podido conservar y, así, han pasado a formar parte de obras de autores posteriores. Durante muchos años la posibilidad de saber en qué medida –en términos cuantitativos– un autor hacía uso de obras pertenecientes a autores anteriores y de qué manera las empleaba –en términos cualitativos– resultaba una labor ardua y prolongada, en ocasiones imposible de descifrar.



Por este motivo, la premisa desde la que parte este proyecto es tan esperanzadora, pues parece que los resultados que se han obtenido hasta ahora ofrecen grandes expectativas, como puede apreciarse en las demostraciones habilitadas en la página web del proyecto en la pestaña *Our Pilot*<sup>1</sup>. Las primeras comparaciones intertextuales realizadas con una amplia base de fuentes, todavía en vías de desarrollo y que está abierta a la colaboración online (a través de OpenITI<sup>2</sup>), muestran lo que ya se suponía, esto es, que la tradición escrita árabe medieval es considerablemente intertextual. Así pues, las posibilidades de estudio y comparación de fuentes árabes que se presentan gracias al empleo de esta herramienta se amplían y van a permitir, entre otros aspectos, los siguientes: conocer la evolución sufrida por los textos a lo largo del tiempo y del espacio, detectar el grado de distorsión ideológico que han experimentado, determinar la difusión de una obra de un territorio a otro, estudiar cuestiones de autoría y originalidad, averiguar las formas de transmisión del saber, identificar las adaptaciones textuales y los diferentes formatos que un mismo texto u obra podían presentar, etc.; en definitiva, un sinfín de posibilidades que permiten su aplicación en otras tradiciones literarias con un alto grado de intertextualidad.

En relación con la base de datos o corpus, las fuentes han sido divididas geográficamente en las siguientes áreas: Iraq, Península Arábiga y Oriente Islámico, Egipto, Occidente Islámico (al-Ándalus y el Magreb), Siria y territorios anexos; todas ellas datadas desde el siglo VIII hasta finales del siglo XVI. Esta base de datos –cuyo objetivo es ser lo más grande posible para lograr los mejores resultados– se puede consultar en línea<sup>3</sup> y se encuentra, como hemos referido, abierta a la colaboración de quienes dispongan de ejemplares editados de fuentes árabes.

De todos los proyectos actuales, relativos al ámbito de los estudios árabes e islámicos, el Proyecto KITAB es el que está empleando más técnicas de las Humanidades Digitales, a diferencia de otros proyectos que se han quedado en valiosas fuentes de datos, como es el caso del PUA de la Escuela de Estudios Árabes de Granada<sup>4</sup>, o el HATA<sup>5</sup> del Centro de Ciencias Humanas y Sociales de Madrid<sup>6</sup>.

El proyecto KITAB ha creado, para la comparación de textos, un software de reutilización textual (en inglés *text reuse*) llamado *passim*<sup>7</sup>, basado en un algoritmo desarrollado por David Smith de la Universidad de Northeastern. Este algoritmo fue, en un principio, escrito

<sup>1</sup> Accesible desde: <http://kitab-project.org/our-pilot/>.

<sup>2</sup> Accesible desde: <http://kitab-project.org/openiti/>.

<sup>3</sup> Accesible desde: <http://iti-corpus.github.io/> y <http://github.com/OpenITI>.

<sup>4</sup> Accesible desde: <https://www.eea.csic.es/pua/>.

<sup>5</sup> Accesible desde: <http://kohepocu.cchs.csic.es/>.

<sup>6</sup> Actualmente, la Universidad de Gante está desarrollando una base de datos bajo el proyecto de *Mamluk Prosopography Project* (MPP), cuyos resultados y utilidad apenas conocemos, pero que seguramente serán una gran herramienta para el periodo mameluco islámico. Accesible desde: <https://bit.ly/3ORBW9G>.

<sup>7</sup> Accesible desde: <https://github.com/dasmiaq/passim>.

para estudiar la presencia de los llamados textos virales que circulaban a través de los textos periodísticos americanos del siglo XIX existentes en la Biblioteca del Congreso de los Estados Unidos<sup>8</sup>. Actualmente, lo han adaptado para poder aplicarlo en un corpus más amplio y en caracteres árabes. Este software, a grandes rasgos, realiza una comparación entre las palabras (con base en una unidad de medida variable, normalmente 5 palabras por unidad) contenidas en un texto y las de otro u otros textos. Este procedimiento comparativo se repite incontablemente, elaborando, cada vez que encuentra una coincidencia, un registro incluido en una base de datos. Según la página web del proyecto, el primer estudio piloto comparativo incluyó 3.500 libros que dieron como resultado 650.000 coincidencias o registros, lo que evidencia que, al menos, cuando dos libros se comparan, se genera como mínimo un registro o coincidencia entre ambos. El procedimiento y la explicación técnica del funcionamiento de *passim*, en cuyo desarrollo ha contribuido notablemente el profesor Maxim Romanov de la Universidad de Leipzig, está perfectamente explicado en la pestaña *Our Pilot* de la página web del presente proyecto, donde se ofrecen datos precisos del proceso de comparación y de las medidas de unidad empleadas para ello.

Finalmente, en lo que respecta a dicha página web (cuyo contenido se presenta en inglés), cabe decir que es bastante clara y accesible, incluso para quienes no estén muy familiarizados con las fuentes árabes e islámicas. Asimismo, gran parte de la información que ofrece está relacionada con las técnicas digitales y computacionales que han desarrollado y pretenden desarrollar, así como con las aplicaciones que desean emplear en las mismas, tales como extraer datos de otra índole, como los relativos a grupos de pasajes coincidentes en los libros. En este caso, el algoritmo genera un índice a partir de los pasajes de las unidades repetidas y, posteriormente, extrae pares de documentos de este repertorio generado. Según los especialistas, este segundo tipo de datos será especialmente útil para estudiar las redes a través de las cuales se han transmitido las ideas.

Actualmente, a través de esta iniciativa, los investigadores e historiadores en el campo de las Ciencias Sociales y los estudios islámicos podemos disponer de una herramienta digital que nos permite rastrear una obra o fragmento de la misma en obras posteriores, así como conocer en qué medida un autor o autores copiaron en sus obras fragmentos de obras previas, además del uso que hicieron de ellas. Esta revelación que, hasta hace poco, era algo inconcebible ahora podrá lograrse gracias a labor realizada por los investigadores de este proyecto KITAB.

---

<sup>8</sup> En total, analizó 1,6 billones de palabras de 132 periódicos.