

DirecciónClara **Martínez
Cantón**Gimena **del Río
Riande**Francisco **Barrón****Editor Asociado**Rubén **Íñiguez
Pérez****Al-Maktaba al-Shamila**<https://shamela.ws/>

Reseña realizada por:

Covadonga BARATECH SORIANO

covadonga.baratech@cchs.csic.es

Consejo Superior de Investigaciones Científicas (CSIC)

<https://orcid.org/0009-0008-8337-0141>

En 1949 comenzó el proyecto *Index Thomisticus* del jesuita italiano Roberto Busa, que, junto a *Computerized Concordance to the Revised Standard Version of the Bible* del cura John W. Ellison (1957), dieron el pistoletazo de salida a las Humanidades Digitales. Ambos proyectos partían de la creación de un corpus de textos (las obras de Santo Tomás de Aquino en el caso de Busa, la Biblia en el de Ellison) que fue tratado y editado para permitir su análisis: Roberto Busa buscaba realizar una lematización de toda la obra del filósofo mientras que John W. Ellison trabajaba sobre las concordancias. Así, la creación de corpus está en el origen mismo de las Humanidades Digitales.

Aunque no ha pasado tanto tiempo desde entonces (74 años desde el *Index Thomisticus*), el desarrollo de las Humanidades Digitales ha sido abrumador. A estas alturas, los corpus digitales son algo que se da por sentado entre los que trabajan con ciertos alfabetos, pero resultan relativamente nuevos para otros. Este es el caso en los estudios árabes e islámicos (o “islamicados”, “islamicate” en inglés, término acuñado por Marshall Hodgson (1922-1968) para intentar distinguir entre lo directamente relacionado con la religión islámica y lo relacionado con una cultura en la que el islam juega un papel relevante)¹. Dentro de los estudios árabes e islámicos, la compilación de corpus es un hito reciente, pues empezó a darse en la primera década de los 2000. La mayoría de los corpus que han aparecido desde entonces buscan estudiar el árabe moderno estándar y están compuestos principalmente de textos provenientes de

¹ Para más información sobre el debate del término “islamicado”, véase Morrissey, F. (9 de julio de 2021). *The Venture of the Islamicate: The History of a Key Concept in Islamic Intellectual History*. [Entrada de blog]. <https://intellectualhistory.web.ox.ac.uk/article/venture-islamicate-history-key-concept-islamic-intellectual-histor>

prensa escrita online. También destaca el desarrollo en los últimos años de corpus para el estudio del árabe dialectal. En cuanto al estudio diacrónico de la lengua árabe, los corpus creados para su efecto son escasos. The King Saud University Corpus of Classical Arabic (KSUCCA) de 2013, Elewa (2004) y Tashkeela son algunos ejemplos de corpus creados para el estudio del árabe clásico medieval, mientras que corpus como KACST Arabic Corpus (Al-Thubaity, 2015), Leeds Arabic Internet Corpus², International Corpus of Arabic³, Arabi Corpus⁴ y CLAUDia incluyen el árabe medieval pero no se limitan a él. Por desgracia, muchos de estos corpus no son accesibles ni públicos, como es el caso del corpus Elewa⁵.

Varios de estos corpus extraen sus textos del repositorio online Al-Maktaba al-Shamila (“La biblioteca completa”). Con más de 8000 obras de 3000 autores⁶, Al-Maktaba al-Shamila es un proyecto que busca poner al alcance de todo el mundo y de forma gratuita todos los libros relacionados con lo que se denominan las ciencias religiosas (“al-‘ulūm al-shar‘iyya” en árabe). No obstante, la acepción tradicional de ciencias religiosas es muy amplia, por lo que Al-Maktaba al-Shamila incluye obras historiográficas, filológicas, literatura de viaje, gramáticas y textos médicos, entre otros. Este interés concreto por los textos de las ciencias religiosas nace del hecho de que el proyecto está patrocinado por una asociación caritativa, aunque ahora mismo cualquiera puede apoyar Al-Maktaba al-Shamila a través de su Patreon⁷. Como puede verse a través de esta plataforma de micromecenazgo, Al-Maktaba al-Shamila crea e-books, textos digitalizados diseñados para leerse de forma individualizada, y no para funcionar como un corpus. La mayoría de los libros disponibles en Al-Maktaba al-Shamila fueron digitalizados y revisados de forma manual, pero para unos pocos el proceso fue automatizado (y así queda registrado en los metadatos de la web). Una de las grandes ventajas de Al-Maktaba al-Shamila es que incluye los datos de los textos impresos digitalizados, como la editorial, el año de publicación o el número de edición, lo que permite a los investigadores citar con precisión las obras⁸.

Al-Maktaba al-Shamila puede consultarse online o puede descargarse. En la portada de la página web hay tres enlaces. El primero lleva al catálogo de temas, en el que se indica el número de obras que incluye cada apartado. Los textos de al-Shamila están clasificados en cuarenta apartados que abarcan desde obras de doctrina (“‘aqīda”) hasta obras de retórica (“balāgha”). Al entrar en los apartados encontramos la lista de todos los libros disponibles en cada tema. Bajo el

² Accesible desde: <http://corpus.leeds.ac.uk/internet.html>, por el momento no accesible sin usuario de la universidad de Leeds, por lo que se puede consultar aquí: <https://www.latl.leeds.ac.uk/resources/corpora-and-corpus-tools/>.

³ Accesible desde: <http://www.bibalex.org/ica/en/About.aspx>.

⁴ Accesible desde: <https://arabicorpus.byu.edu/>.

⁵ Véase Belinkov, Y., Magidow, A., Romanov, M., Shmidman, A. y Koppel, M. (2016). Shamela: A Large-Scale Historical Arabic Corpus (p. 46). En E. Hinrichs, M. Hinrichs y T. Trippel (Eds.). *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)* (pp. 45-53). The COLING 2016 Organizing Committee.

⁶ A marzo de 2023.

⁷ Accesible desde: <https://www.patreon.com/shamela4>.

⁸ Véase Belinkov, Y., Magidow, A., Romanov, M., Shmidman, A. y Koppel, M. (2016). Shamela: A Large-Scale Historical Arabic Corpus (p. 47). En E. Hinrichs, M. Hinrichs y T. Trippel (Eds.). *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)* (pp. 45-53). The COLING 2016 Organizing Committee.

título de la obra y el nombre de su autor, hay un breve resumen que indica de qué versa, estos resúmenes pueden ocultarse y ver solo los títulos de las obras y sus autores. En cada apartado hay un buscador de obras. El segundo enlace dirige al panel de búsqueda, donde se podrá buscar por contenido, por nombre de libro y por nombre de autor. Las búsquedas por nombre de libro y nombre de autor serán siempre simples y han de ser exactas para que arrojen resultados, mientras que las búsquedas por contenido podrán ser simples o avanzadas. Además, las búsquedas simples por contenido admiten el uso de símbolos como las comillas, los de suma y resta y los paréntesis para hacer búsquedas más complejas. Por último, el tercer enlace conduce a las últimas obras añadidas, donde podemos ver el título de la obra, su autor y un breve resumen de su contenido.

En la pestaña del índice de autores también puede buscarse a un autor concreto o se lo puede localizar de forma manual en los tres índices que existen: el primero se organiza por fecha de muerte, el segundo por orden alfabético y el tercero por número de obras producidas. Por último, está la pestaña desde la que se puede descargar Al-Maktaba al-Shamila. La biblioteca está disponible tanto para dispositivos móviles (ya tengan Android como sistema operativo o iOS) como para ordenadores, en esta última opción puede descargarse la biblioteca entera o parte de ella. Si se opta por la descarga parcial, podrán añadirse nuevas obras a la descarga más adelante. Actualmente Al-Maktaba al-Shamila solo funciona en sistemas operativos Windows, desde la versión 7 en adelante, aunque esperan sacar una versión para Linux y para Mac próximamente.

Una de las ventajas de la aplicación de Al-Maktaba al-Shamila es que permite acceder a los textos incluso cuando no se cuenta con acceso a internet, pero no es la principal. La principal ventaja de la aplicación de Al-Maktaba al-Shamila es que ofrece la posibilidad de descargar los textos íntegros. Sin embargo, los textos se descargan en formato BOK y deberán ser convertidos a formato PDF, EPUB o MOBI para su uso posterior⁹. Esta opción facilita enormemente tanto la lectura del texto como su tratamiento con otras herramientas propias de las Humanidades Digitales y no es posible desde la página web, donde lo máximo que se puede hacer es copiar de forma manual página por página el texto de la obra, lo cual dificulta su tratamiento posterior.

La segunda gran ventaja de la aplicación con respecto a la versión web es la complejidad de las búsquedas que se pueden realizar. Gracias a las distintas opciones de búsqueda se puede eliminar las obras de las cuales no hay copias completas y solo quedan fragmentos, así como los textos que nunca se han editado y que solo existen en copias manuscritas, entre otras posibilidades. En general, en la aplicación se accede a la información de forma más rápida e intuitiva. Por ejemplo, la primera pantalla ofrece una lista completa de todas las obras disponibles junto a su autor, ordenadas por la fecha de fallecimiento de este. Al posar el cursor por su nombre, el programa muestra una ficha del autor que incluye su nombre completo, fechas de nacimiento y muerte y una breve biografía suya. Además, se pueden consultar fácilmente el apartado al que pertene-

⁹ En shamela.org existe un conversor de formatos que incluye además un apartado con libros ya convertidos previamente que se pueden descargar de forma gratuita. Accesible desde: <https://shamela.org/>.

cen las obras, la editorial, la edición y el año de publicación. Otra posibilidad que ofrece la aplicación, y que no está presente en la página web, es que incluye el Corán, en el que se pueden hacer búsquedas de idéntica complejidad que en el resto de las obras. Por último, existe la opción de guardar textos favoritos y de descargar nuevos textos.

Al-Maktaba al-Shamila es una herramienta fundamental para los investigadores de las sociedades islámicas medievales, sobre todo para historiadores y especialistas en derecho y religión, pues permite un acceso fácil y gratuito a un enorme corpus de textos sobre el cual puede realizarse búsquedas de gran complejidad. No obstante, debido a que los textos pertenecen a las ciencias religiosas, la biblioteca no incluye un gran número de obras clásicas sobre literatura, filosofía o ciencias empíricas. En cuanto a los investigadores que centran sus estudios en el idioma árabe, tienen a su disposición una gran fuente para sus investigaciones, pero estos textos han de ser tratados para que resulten útiles. Un ejemplo de esto, además de varios de los corpus mencionados previamente, es el Open Islamicate Texts Initiative (OpenITI) de Sarah Bowen Savant, Maxim Romanov y Mathew Thomas Miller⁸.

⁸ El proyecto Open Islamicate Texts Initiative (OpenITI) fue reseñado por Abdenour Padillo (véase Padillo, A. (2019). Reseña de Open Islamicate Texts Initiative (OpenITI). *Revista de Humanidades Digitales*, 4, 187-191. <https://doi.org/10.5944/rhd.vol.4.2019.2416>). Para saber más sobre el tratamiento de los textos extraídos de Al-Maktaba al-Shamela véase los procedimientos del Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH) de Osaka, Japón, presentado por los directores del proyecto.