

## Otros informes

### El peso digital de la tradición clásica: de Sumeria a Roma

Josep Turiel, CRAI Biblioteca de Lletres, Universitat de Barcelona



Los estudios clásicos tienen una particularidad: la interdisciplinariedad. Tratan de filología, arte, epigrafía, filosofía, arqueología, historiografía, etc. En ellos hay que interrelacionar disciplinas y de acertar las tecnologías apropiadas en la recuperación, clasificación, etiquetado, orden, presentación, lectura (OCR) y visualización de datos. A

ello se suman aspectos tan importantes como la fragmentación de la herencia clásica (qué nos ha llegado y cómo), las diferentes copias de los originales revisadas o la reconstrucción hecha a partir de lo que se conoce de un texto original.

El trabajo que reseñamos tiene una amplia introducción a la historia reciente de la digitalización del patrimonio clásico pero fundamentalmente es un estado de la cuestión, evitando las habitualmente largas, enciclopédicas y aburridas listas de herramientas y recursos, matizando mucho los aspectos tecnológicos y aportando soluciones de futuro para la suma ingente de proyectos nacidos en los últimos 40 años. Está pensado en primer lugar para los profesionales de la digitalización

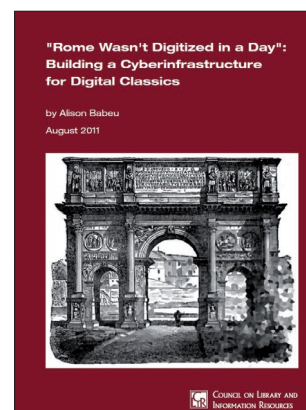
y en segundo lugar investigadores del mundo clásico, ofreciendo pautas para gestionar recursos de calidad del mundo clásico, para hacer una sesión de formación para arqueólogos, etc.

El informe repasa cronológicamente algunos de los grandes proyectos digitales que, más o menos fragmentarios, se limitaban en una primera fase a textos disponibles online.

Una 2ª generación formada por los grandes corpus (como el del centro *Thesaurus Linguae Graecae*, *TLG*, de la *UCA Irvine*) y el reto de la captura de otros alfabetos como el griego o el copto. En los años 80 aparecen los primeros proyectos que incluyen los formatos de marcación diseñados por la *Text Encoding Initiative* (los que llevan el proyecto *Perseus*, de la *Tufts University*).

La 4ª generación es la de la digitalización masiva de los años 90 (*Google Books* y la *Open Content Alliance*). Faltaría una 5ª fase que la autora resume como la era del marcado automático, la personalización de los textos o la industrialización del trabajo realizado).

Otros apartados tratan del uso que se hace de los proyectos, de la tipología de los usuarios



### Perseus digital library

Perseus es una biblioteca digital en evolución, iniciada en 1985, que actualmente se centra en el mundo griego antiguo. Contiene muchas transcripciones y traducciones de obras griegas y un número creciente de latinas, así como del Renacimiento. Incorpora herramientas de análisis de texto, principalmente para trabajar con los textos griegos. También contiene una gran cantidad de recursos de arte y arqueología, e información general sobre el mundo antiguo. Cuenta con un catálogo de 25.000 imágenes de arquitectura,

escultura, monedas, vasos, y sitios, y una extensa colección de atlas, mapas y planos.

Perseus es un proyecto de la *Text Encoding Initiative (TEI)*, *Classics Department, Tufts University*, Medford, Massachusetts, y es un referente básico.

<http://www.perseus.tufts.edu/hopper/>

<http://www.tei-c.org/Activities/Projects/pe01.xml>

habituales, y de los hábitos de investigación en el ámbito humanista. Destaca la autora como es de desalentador en pleno siglo XXI el desconocimiento que sobre el uso de las TIC tienen algunos investigadores de humanidades.

turielj@ub.edu

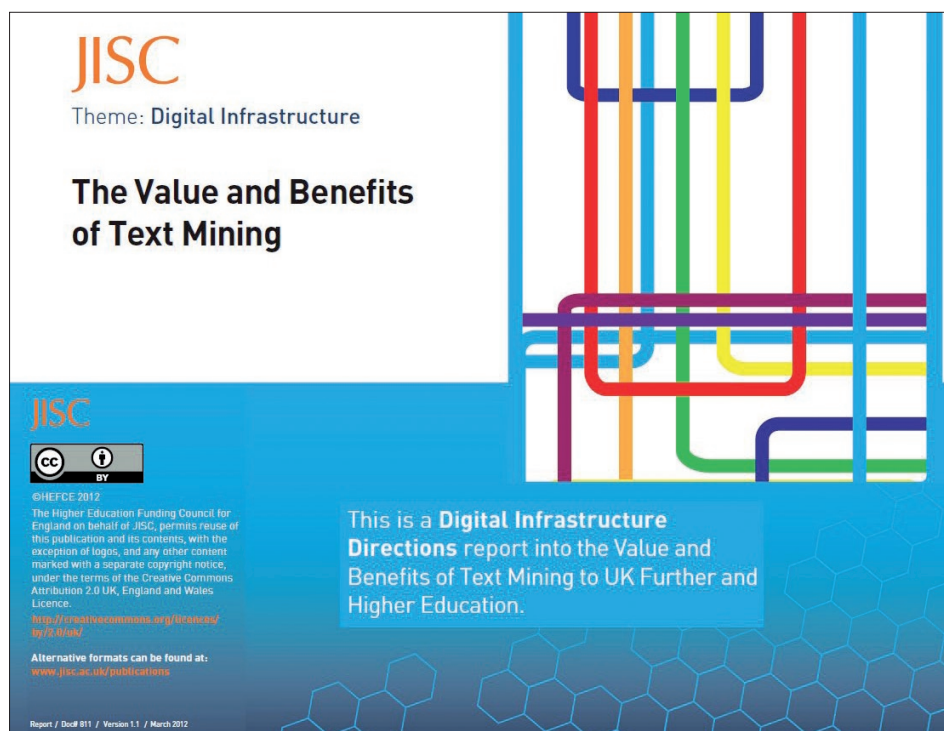
---

**Babeu, Alison.** *Rome wasn't digitized in a day: building a cyberinfrastructure for digital classicists.* CLIR Activities, Council on Library and Information Resources, August 2011, 307 pp.  
<http://www.clir.org/pubs/abstract/pub150abst.html>

---

Nota original publicada en *Blok de BiD*:  
<http://www.ub.edu/blokdebid/es/content/el-peso-digital-de-la-tradición-clásica-de-sumeria-roma>

## Valor y beneficios de la minería de textos



Informe sobre las barreras al sector del *text mining*, relacionado en parte con el big data, vigilancia tecnológica, recuperación de información y las plataformas de *social listening*.

Todos los días se genera una gran cantidad de nueva información y de datos debidos a las actividades económicas, académicas y sociales. Se prevé que este mar de datos, que tiene un importante valor potencial económico y social, aumentará a

razón del 40% anual. Para aprovecharlo se usan técnicas tales como la minería y el análisis de textos y datos.

Las empresas utilizan estas técnicas para analizar los datos de los clientes y la competencia y así mejorar la competitividad; la industria farmacéutica hace minería en las patentes y los artículos de investigación para descubrir fármacos; dentro de la investigación académica, la minería y análisis de grandes conjuntos de datos optimiza el ahorro y ayuda a obtener nuevos conocimientos en áreas tan diversas como biología, física de partículas, medios de comunicación o comunicaciones.

La comunidad global de investigación genera más de 1,5 millones de artículos nuevos académicos al año<sup>1</sup>. Tal como el informe de **Hargreaves**<sup>2</sup> puso de relieve, el análisis y la minería de textos de esta literatura académica y de otros textos digitalizados ofrece una oportunidad real de apoyo a la innovación y al desarrollo de nuevos conocimientos.

Sin embargo, las actuales leyes de copyright del Reino Unido restringen la minería de textos. Para remediarlo, **Hargreaves** propone una excepción que permita hacerlo en casos de investigación no comercial.

Para ser "minado", el texto debe ser accesible, copiado, analizado, comentado, y relacionado con la información y el conocimiento existentes. Incluso si el usuario tiene derecho de acceso a las obras, la realización de copias anotadas puede ser ilegal.

### Notas

1. **Ware, Mark; Mabe, Michael.** *The stm report: An overview of scientific and scholarly journal publishing*, 2009.

<http://www.stm-assoc.org/document-library>

2. **Hargreaves, Ian.** *Digital opportunity: a review of intellectual property and growth*, 2011.

<http://www.ipo.gov.uk/lipreview-finalreport.pdf>

---

**McDonald, Diane; Kelly, Ursula.** *Value and benefits of text mining.* JISC, 2011, 36 pp., 6,8 MB

<http://www.jisc.ac.uk/publications/reports/2012/value-and-benefits-of-text-mining.aspx>

---