Liberate the Text!
in 18thConnect and TypeWright
www.18thconnect.org/typewright

TypeWright allows crowd-sourced correction of Optical Character Recognition (OCR) generated text for early modern printed documents, specifically those in Gale/Cengage Learning’s Eighteenth-Century Collections Online (ECCO) subscription database and, soon, Proquest’s Early English Books Online (EEBO). The current OCR output for ECCO is not accurate enough for full-text searching, data mining, and the creation of digital editions that meet the standards required of scholarly work.

By the end of 2014, TypeWright will contain the 45 million page images and text that make up this combined collection of the primary, English-language, digitized source material for 15th-18th Century Humanities scholarship.

At 18thconnect.org, users search for “TypeWright enabled” texts (currently 183,000 ECCO documents).

Clicking “edit” ports directly into the editing interface, which displays snippets of the page image.

Lines on the page image correspond to OCR output in the text editing box, which can be corrected or confirmed.

A history of user edits is available for collaboration and project planning.

Corrected text, per our agreements with contributing databases, return to these products to make our cultural heritage fully searchable and visible.

Users that liberate the text use these liberated materials for digital projects - which can be submitted to 18thConnect for peer review and aggregation.

After a document is corrected, our TypeWright team reviews the work via criteria based on common OCR errors.

If the work passes the evaluation process, the text can be liberated! - users receive the corrected plain text or XML/TEI-encoded files of their transcription.