

Text Digitization of Library Materials: OCR A Brief Bibliography

Reports

Bicknese, Douglas A. [*Measuring the Accuracy of the OCR in the Making of America*](#). Ann Arbor, Mich.: University of Michigan Library, 1998.

Report of Michigan's investigation to correlate character accuracy with Prime Scores (generated by [Prime Recognition OCR software](#)) to determine whether uncorrected OCR would adequately meet the indexing and retrieval requirements for the [Making of America](#) collection.

Library Digital Initiative Project Team, Harvard University Library. [Measuring Search Retrieval Accuracy of Uncorrected OCR: Findings from the Harvard-Radcliffe Online Historical Reference Shelf Digitization Project](#). Cambridge, Mass.: Harvard University Library, 2001.

Report of Harvard's investigation to evaluate the conditions for search retrieval successes and failures when using uncorrected OCR for indexing. Based upon analysis of findings from the [Harvard-Radcliffe Online Historical Reference Shelf](#) project.

Selected Articles

Fenton, Eileen Gifford and Hoyt N. Duggan. "[Electronic Textual Editing: Effective Methods of Producing Machine-Readable Text from Manuscript and Print Sources](#)." Text Encoding Initiative Consortium, 2002.

Includes section providing "Fundamentals of the OCR Process" for JSTOR (as of 2002), as well as useful "Key Points" to consider to achieve good OCR quality.

Stehno, Birgit, Alexander Egger and Gregor Retti. "[METAe—Autoamted Encoding of Digitized Texts](#)," *Literary & Linguistic Computing* 18(1):77-88, 2003.

Detailed explanation of use of METS to incorporate data structure of the METAe ALTO (Analysed layout and text object) file, which contains layout structures and OCR-generated text output by the METAe engine.

Tanner, Simon. "[Deciding whether Optical Character Recognition is feasible](#)," King's Digital Consultancy Services, December, 2004.

Introduction to character recognition techniques and recommendations for assessing historic materials and other key variables to determine whether OCR will be an effective process.

Available from <http://preserve.harvard.edu>
<http://preserve.harvard.edu/bibliographies/ocr.pdf>

Last updated: February 12, 2007