What's new:
- New tools for lines, words, glyphs
- Integrated OCR (Tesseract 3.02)
- ALTO and FineReader XML support
- Text overlay
- Drag and drop files into Aletheia

Open document

Recent Documents:
- cbl1692_110_002boxfilesglyphsonly.xml
- BRA2788_002.boxfileslyphsonly.xml
- UAGG_quvat_63_09polyqonlyphsonly.xml
- UAGG_quvat_83_09qlyphsonly.xml
Step One: Naming Font Sets

Page images are named according to publisher, publication year, and the 20-line height. For example, a text image from the 1702 Anno Regni text would be named cbil1702_116.

<table>
<thead>
<tr>
<th>Printer</th>
<th>Primary Type Style</th>
<th>20-LN HT (mm)</th>
<th>Body Size</th>
<th>Year</th>
<th>Place</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrew Crook</td>
<td>Roman</td>
<td>143</td>
<td>Double Pica</td>
<td>1651</td>
<td>London</td>
<td>Hobbes</td>
<td>Leviathan</td>
</tr>
<tr>
<td></td>
<td>Roman</td>
<td>92</td>
<td>English</td>
<td>1651</td>
<td>London</td>
<td>Hobbes</td>
<td>Leviathan</td>
</tr>
<tr>
<td></td>
<td>Italic</td>
<td>92</td>
<td>English</td>
<td>1651</td>
<td>London</td>
<td>Hobbes</td>
<td>Leviathan</td>
</tr>
<tr>
<td>Charles Bill</td>
<td>Black Letter</td>
<td>116</td>
<td>Great Primer</td>
<td>1702</td>
<td>London</td>
<td>England and Wales</td>
<td>Anno Regni</td>
</tr>
<tr>
<td></td>
<td>Roman</td>
<td>116</td>
<td>Great Primer</td>
<td>1702</td>
<td>London</td>
<td>England and Wales</td>
<td>Anno Regni</td>
</tr>
<tr>
<td></td>
<td>Black Letter</td>
<td>116</td>
<td>Great Primer</td>
<td>1693</td>
<td>London</td>
<td>England and Wales</td>
<td>Gulielmum et Mariae</td>
</tr>
<tr>
<td></td>
<td>Roman</td>
<td>116</td>
<td>Great Primer</td>
<td>1693</td>
<td>London</td>
<td>England and Wales</td>
<td>Gulielmum et Mariae</td>
</tr>
<tr>
<td></td>
<td>Roman</td>
<td>94</td>
<td>English</td>
<td>1693</td>
<td>London</td>
<td>England and Wales</td>
<td>Gulielmum et Mariae</td>
</tr>
<tr>
<td></td>
<td>Black Letter</td>
<td>116</td>
<td>Great Primer</td>
<td>1693</td>
<td>London</td>
<td>England and Wales</td>
<td>Gulielmum et Mariae</td>
</tr>
</tbody>
</table>
Step Two: Page Segmentation and Analysis

Layout regions, lines, words, and individual glyphs in a text are identified and defined using Aletheia Desktop, a tool developed by PRImA Lab at the University of Salford that performs page segmentation and text recognition, among other tasks.
TO THE MOST REVEREND FATHER in God, and his singular good Lord, Edmund, by the grace of God, Archbishop of Canterbury, Primate of all England, and Metropole of the See of St. Peter, in Rome, &c.

After long exercise in translating such Latin Commentaries upon the holy Scriptures, as I showed myself like to further my countrymen, which understanding I that tongue, to the solid knowledge of true Religion: at last I took in hand Mr. Calvines exposition upon Saint Pauls Epistles to the Corinthians. And, as in my poor judgment, the writer is a most excellent instrument of God, for the simple setting forth of his truth, so in making my choice...
TO THE MOST REVEREND FATHER IN GOD, AND HIS SINNER GOOD LORD, EDMOND, BY THE
GRACE OF GOD, ARCHBISHOP OF CANTERBURY, PRIMATE AND METROPOLITAN OF ALL ENGLAND, THOMAS TIMMUS
WALKER IN THE SALUTARY INSTRUCTIONS OF THE SCRIPTURES IN GOD'S WORK.

[Latin text transcribed]

A long exercise in translating such Latine Commentaries upon the holy Scriptures, as I thought most like to further my countrymen, which under
TO THE MOST REV.

end Father in God, and his sin-

guler good Lord, Edmond, by the

grace of God, Archebishiop of Can-

terbury, Primate and Metropoli-
tan of England, Thomas Tunne,

and of the plentie of the sea

spoken in Christ 1553.

For long exercise in translating such

Latin Commentaries upon the holy

Scriptures, as if

thought most like to

farther my country.
TO THE MOST REVEREND FATHER IN GOD, and his singular good Lord, Edmund, by the grace of God, Archbishops of Canterbury, and Primate of All England, Thomas Linus, Presserver of the Lamb of God, in Christ's holy Temple.

Fears long exercised in translating such Latin Commentaries upon the Holy Scriptures, as I thought best like to further my country.

For long examine in translating such Latine Commentaries upon the holy Scriptures, as if thought most like to further my country men, which under-
Step Three: Text Correction

Aletheia analyzes the page image and generates text for each character. Students are then responsible for fixing any “misreads” by typing in the correct text in the Text Content Box for each corresponding character--paying special attention to unique characters such as long S’s, ligatures, italics, rotunda R’s, suspension marks, and printer’s marks.
To The Most Reverend Father in God Archbishops and Bishops of all England, 

Whose state and station, by the blessing of God, is most excellent, and whose person doth shew forth in all the power and majesty of an Archbishop and Bishop; and whose hands are full of the plentifull riches of the Spirit in Christ Jesus, 

I desire to make known to you, that in translating the Latine Commentaries upon the Scriptures, as ye thought most likely for the further my country, 

Text Correction

To The Most Reverend
Step Four: OCR Training

10-15 page images are processed for each font type and comprise a training set. A Tesseract training tool uses the font set to train the OCR engine to read and recognize that particular font.