

Create an ALTO xml file

Created By: Dave Allen
Created on: 17/05/2019

ALTO (Analyzed Layout and Text Object) is an open XML Schema developed by the EU-funded project called METAe. The standard was initially developed for the description of text OCR and layout information of pages for digitized material. The goal was to describe the layout and

text in a form to be able to reconstruct the original appearance based on the digitized information - similar to the approach of a lossless image saving operation.

there are many options out there to create an ALTO file -- however the open source route is to use tesseract

you will need to install tesseract V4 and a language pack

the command to create a ALTO file is as follows

```
tesseract /permanent_storage/archive/images/slq/pub/2019-05-14/archive/690444-v012n006/  
690444-v012n006-s0002.tif output-filename -l eng alto
```

Hope you find this article of value

[Report](#)