

## Historic Pittsburgh Full-Text: Running PrimeOCR Job Server

---

### Some explanation of the OCR application first:

The OCR engine uses files called “**jobs**” that tell it what to do. The job file is a text file containing

1. The path to the directory or specific image file that should be processed
2. The path to the “template” file for that image.

The “**template**” file contains a series of co-ordinates that restrict the OCR engine to certain parts of the image. These co-ordinates are created by using the PRView application to pre-zone an image file.

You only need to tell the OCR engine where to look for the job files, because the job files then point to the image files and the templates to use when processing them.

When the OCR server runs, it will work on all job files that it finds in the designated place. It will process jobs in order of their file-modification date, from oldest to newest.

### Now, the Steps in OCR-ing a Batch:

At this point, all of the pre-zoning jobs should have been completed for the batch. These must be run through the OCR process first. Afterwards, one final job will be created that will OCR all of the images that were not covered by pre-zoning.

1. Open Prime Job Server
2. Under *Settings*:
  - Make sure that the path to the job directory is correct
  - Choose ‘delete the job file when completed’
  - **DO NOT** continuously poll the job directories
  - Choose recover from errors
  - **DO NOT** set it to delete the image files
  - Save the files to the same directory as the images

Click “Start” to begin processing the pre-zoned jobs. The Job Server will take a while to first count all of the jobs that it has to perform, and will then begin to work on jobs in chronological order based on the file’s modification date.

When the pre-zoned jobs are complete, you can use the built-in “Wizard” to create a job that will OCR the rest of the images.

**IMPORTANT:** Make sure that the output settings **do not** overwrite the OCR output that was created in the previous step. There is a setting choice to “skip” images for which output already exists.

### 3. Select *Wizard*

- Choose “create a job file only”
- Select *Process all image files in a directory and subdirectories.*
- Give full path of the directory you want to process.
- Select *Template*
- Choose 5engine2.ptm as your template
- Click on Finish
- Name your Job
- Exit

4. At the Job Server select *Start*. The OCR will take a while so let it run.

### 5. Dealing with Errors

After the OCR process is finished any errors will be in red. To troubleshoot errors check the Log to determine what files had problems. There are no obvious or common problems that occur but one thing to check is that all files have corresponding Jobs and Templates. Go back to preview and open the offending file. Create a new job and template, **only** use a unique name. **Do not overwrite** existing jobs and templates. Use extra letters (ie: replace 0004000d with xx0004000d) The Jobs should be located in `p:\ocr\viewjob`.