College Board Technical White Paper—
Batch Essay Delivery

Using an index file to extract student SAT essay images for upload into an admission system or database

This document provides information and recommendations to aid colleges and universities in developing a mechanism to extract student SAT essay images from batch essay PDF files and upload the images directly into their admission system or database for review in the admission process.

Description of the Batch Essay Process

Institutions can choose to receive all student essays in batch files. Once this option has been selected, and when batches of essays are available, you will receive an email notification.

Each batch file may contain as many as 100 to 500 student essays (depending on your preferred file size). The number of files you receive will depend on the number of students who sent scores to your institution—as many PDFs as necessary will be sent to incorporate the entire output of student essays.

Description of the Index File

A complete index file accompanies each batch of essays. The index file lists which students' essays can be found in each PDF file, and can be used to upload essay images automatically into a database. This index file can be used to help program the automatic extraction of images. This is a fixed-width ASCII format file, as specified below.

The name of the index file is unique to each batch of essays, and adheres to the following format: **complete_index_<BB>.txt** (where BB is the 2-digit batch number).

Technical Information on Using the Index File

To extract and load essay images, you will need a tool that allows you to manipulate PDF files. Below is a *partial* list of tools that may be used for this purpose.

- **JAVA:** http://www.lowagie.com/iText/
- **.NET:** https://sourceforge.net/projects/itextsharp/
- **VB/COM:** http://www.devdirect.com/All/PDFCREATE_PCAT_1876.aspx

Once you have installed a tool to allow you to extract a PDF into your admissions system or database, follow these steps:

- Open the *batch file name* referenced in Field 6.
- Go to the *batch file page number* referenced by Field 7.
- Extract that image using a tool like one of those mentioned above.

The following table specified the fields and field lengths contained in the index file created for each batch. **Note:** Each row in the file represents one student essay image.

Field Name	Start Position	End Position	Length	Description
1. Essay Locator ID	1	10	10	The Essay Locator ID of the student's essay containing the image.
2. Student First Name	11	25	15	The first, middle, and last (family) name of the student.
3. Student Middle Initial	26	26	1	
4. Student Last Name	27	41	15	
5. Essay Image Page Number	42	51	10	The **page of the student's essay.** Most essays are 2 pages in length so each student will have 2 entries; some essays, written on large-block answer sheets, are 4 pages in length, and the student will have 4 entries.
6. Batch File Name	52	71	30	The actual file name of the PDF that contains the essay image. The name of the file follows the following format: **Essay_<DDDD>_<BB>_<SSS>.pdf**, (where DDDD is the 4-digit college code, BB is the 2-digit batch number [begins anew with 01 each administration year], and SSS is the 3-digit sequence number [begins with 001]).
7. Batch File Page Number	82	91	10	**The page number in the PDF file** referenced by the batch file name, which contains the image of the student's essay page.

Sample Index File

Below is a sample index file:

```
CTH777PY  EMIL      ESLATE      1    ESSAY_6609_13_1.pdf   1
CTH777PY  EMIL      ESLATE      2    ESSAY_6609_13_1.pdf   2
PLZ312WA  ASHLEY    RSODICH     1    ESSAY_6609_13_1.pdf   3
PLZ312WA  ASHLEY    RSODICH     2    ESSAY_6609_13_1.pdf   4
AHW383FT  VY        AWINTERS    1    ESSAY_6609_13_1.pdf   5
AHW383FT  VY        AWINTERS    2    ESSAY_6609_13_1.pdf   6
```

Essay Locator ID

Student Name

Essay Image Page Number

Batch File Name

Batch File Page Number