IDPL-PFOD2: A New Large-Scale Dataset for Printed Farsi Optical Character Recognition

This Google Drive folder contains IDPL-PFOD2 that is an image dataset of printed Farsi text for Farsi optical character recognition researches.

Due to limited storage space in my Google Drive, the test set is available in a separate link:

https://drive.google.com/file/d/17ZBP5FwLY2-2yQkd2NpLSbvP_v5BRThm/view?usp=sharing

while the train and validation sets are available in my Google Drive:

https://drive.google.com/drive/folders/1_X2XPDvQZLPYJxSiaWSF8AceOUWFnWJU?usp=sharing

- **IDPL-PFOD Description**

1. Is an artificial image dataset of printed Farsi text.

2. Has 2003541 images in .png format, and each image contains a line of real Farsi text.

3. The dimensions of the images are 300 * 50 pixels.

We hope that this dataset will be useful to researchers and developers interested in Persian OCR technology and natural language processing applications. If you have any questions or feedback about the dataset, please feel free to contact us at [fatemeh.asadi@eng.uk.ac.ir].