HOWTO: Extract text position using OcrPage and Recognize()

When you recieve a OcrPage from the Recognize() method of your OcrEngine, it contains a collection of OcrRegions. There are two types of Region: OcrTextRegion, and OcrImageRegion. The OcrTextRegion object contains all the information pertaining to the text in the image. To get the OcrTextRegion one must traverse the OcrRegionCollection and cast the OcrRegion as a OcrTextRegion. The standard way to accomplish this is:

```
oreach (OcrRegion r in page.Regions) { OcrTextRegion textregion = r as OcrTextRegion;
if(r!=null) { //use the textregion in here } }
```

Once you have the OcrTextRegion you can traverse the OcrLineCollection contained in OcrTextRegion.Lines. With an OcrLine object you can traverse the OcrWordCollection contained in OcrLine.Words. With an OcrWord object you can traverse the OcrGlyphCollection contained in OcrWord.Glyphs.

There is a project attached to this article contained in SimpleOCRTextRewriter.zip. It takes in an image, Ocrs the image, and then creates a new image with the text printed using our canvas object and the bounds of the Ocr objects.

Original Article:

Q10234 - HOWTO: Extract text position using OcrPage and Recognize()

Atalasoft Knowledge Base
https://www.atalasoft.com/kb2/KB/50231/HOWTO-Extract-text-position-using-Oc...