

کتابخانه iTextSharp کتابخانه‌ای بسیار قوی و معروف، جهت کار با فایل‌های pdf می‌باشد. کارهایی از قبیل ساخت و ویرایش و تبدیل فایل‌های pdf با این کتابخانه قدرتمند به راحتی محیا است. گاهی نیاز داریم تا اطلاعاتی را از داخل یک فایل pdf استخراج کنیم. ما الان می‌خواهیم کل تصاویر موجود در یک فایل pdf را استخراج کنیم. ابتدا باید فایل pdf را باز کنیم. با استفاده از کلاس PdfReader اینکار امکان پذیر می‌شود. سپس باید تعداد کل صفحات را خوانده و بر روی آنها تکرار کنیم و اطلاعات مورد نیاز را پردازش کنیم.

```
var pdfReader = new PdfReader(pdfFile);
for (var pageNumber = 1; pageNumber <= pdfReader.NumberOfPages; pageNumber++)
{
    var pdf = new PdfReader(pdfFile);
    var pg = pdf.GetPageN(pageNumber);
    // process page
}
```

اکنون ما در حلقه به صفحات دسترسی داریم و به ازای هر صفحه لیست منابع توکار آن صفحه را دریافت می‌کنیم:

```
var pdf = new PdfReader(pdfFile);
var pg = pdf.GetPageN(pageNumber);
var res = (PdfDictionary)PdfReader.GetPdfObject(pg.Get(PdfName.RESOURCES));
var xobj = (PdfDictionary)PdfReader.GetPdfObject(res.Get(PdfName.XOBJECT));
if (xobj == null) continue;
foreach (var name in xobj.Keys)
{
    var obj = xobj.Get(name);
    if (obj.IsIndirect())
    {
        // process obj ...
    }
}
```

حالا باید آبجکتهایی که تصویر هستند را به image تبدیل کنیم .

```
var tg = (PdfDictionary)PdfReader.GetPdfObject(obj);
var width = tg.Get(PdfName.WIDTH).ToString();
var height = tg.Get(PdfName.HEIGHT).ToString();
ImageRenderInfo imgRi = ImageRenderInfo.CreateForXObject(new Matrix(float.Parse(width),
float.Parse(height)), (PRIndirectReference)obj, tg);
```

حال یک ImageRenderInfo داریم که باید در دیسک ذخیره کنیم. این کار را در یک متد مجزا انجام می‌دهیم.

```
private void RenderImage(ImageRenderInfo renderInfo, string imgPath)
{
    var image = renderInfo.GetImage();
    using (var dotnetImg = image.GetDrawingImage())
    {
        if (dotnetImg == null) return;
        using (var ms = new MemoryStream())
        {
            dotnetImg.Save(ms, ImageFormat.Jpeg);
            var d = new Bitmap(dotnetImg);
            d.Save(imgPath);
        }
    }
}
```

کد نهایی به صورت زیر می‌باشد:

```
private void ExtractImage(string pdfFile, string imgPath)
{
    var fileCounter = 0;
    var pdfReader = new PdfReader(pdfFile);
    for (var pageNumber = 1; pageNumber <= pdfReader.NumberOfPages; pageNumber++)
    {
        var pdf = new PdfReader(pdfFile);
        var pg = pdf.GetPageN(pageNumber);
        var res = (PdfDictionary)PdfReader.GetPdfObject(pg.Get(PdfName.RESOURCES));
        var xobj = (PdfDictionary)PdfReader.GetPdfObject(res.Get(PdfName.XOBJECT));
        if (xobj == null) continue;
        foreach (var name in xobj.Keys)
        {
            var obj = xobj.Get(name);
            if (obj.IsIndirect())
            {
                var tg = (PdfDictionary)PdfReader.GetPdfObject(obj);
                var width = tg.Get(PdfName.WIDTH).ToString();
                var height = tg.Get(PdfName.HEIGHT).ToString();
                var imgRi = ImageRenderInfo.CreateForXObject(new Matrix(float.Parse(width),
float.Parse(height)), (PRIndirectReference)obj, tg);
                fileCounter++;
                RenderImage(imgRi, imgPath + fileCounter + ".jpeg");
            }
        }
    }
}
```

و به شکل زیر آن را فراخوانی میکنیم.

```
var path = @"C:\";
var fileName = "1.pdf";

var outputPath = path + fileName + @"_extractedImgs\";
Directory.CreateDirectory(outputPath);

ExtractImage(path + fileName, outputPath);
```