OCR with LEADTOOLS:
The Better Choice

1 Introduction

Optical Character Recognition (OCR) is a technology and concept that is familiar to the majority of programmers: take a picture with words and convert it to text. It sure sounds simple, but implementing it well is often much harder than it looks. Much like watching a professional surfer and trying to do it yourself, developers get bruised, tired, and nearly drown in the endless waves of images with varying fonts, bad scans, dust speckles, and paper crinkles finding new problems in your algorithm.

Save yourself some headache and use LEADTOOLS, the most accurate, fast, and easy-to-use OCR SDK on the market! With over twenty-six years of programming experience, a powerful and extensive set of document image cleanup functions, thread-safe OCR for over forty languages, and the time and resources to test millions of images, LEADTOOLS has earned the trust of Fortune 500 companies and individual contractors alike.

Programming with LEADTOOLS couldn't be easier with high-level interfaces that can convert an image to a searchable PDF in only three lines of code. For those who need additional control or wish to do more advanced tasks such as using zones to read words and characters from specific sections of a form, LEADTOOLS provides low-level control over every aspect of your OCR application.

Key Features in LEADTOOLS OCR SDKs

- Fast and Accurate OCR with multithreaded support
- Broad OCR language character set support including Latin, Cyrillic, East Asian, and Arabic
- Save OCR results to over 40 output formats including searchable PDF, PDF/A, Word, and XML
- Full-page and zonal OCR
- Built-in and custom spelling dictionaries to improve OCR results
- Powerful document image cleanup and preprocessing functions to improve OCR results of scanned images
- Native, full-featured OCR libraries for many platforms including desktops, servers, and mobile phones

2

The OCR Code

One of the most important characteristics of any SDK is ease of use. This is a foundational concept for the developers of LEADTOOLS. Here you can see how to convert an image to a searchable PDF in only three lines of code:

```
IOcrEngine ocrEngine = OcrEngineManager.CreateEngine(
    OcrEngineType.Advantage, false);
ocrEngine.Startup(null, null, null, null);
ocrEngine.AutoRecognizeManager.Run(_strInputFile, _strOutputFile,
    DocumentFormat.Pdf, null, null);
```

The null parameters are for classes you can use to customize the output and processing such as file format settings, document cleanup, callbacks, and more. Passing null uses default values which are optimized for the majority of scanned documents. Here's a screenshot of the original TIFF image and searchable PDF created by this code:

3
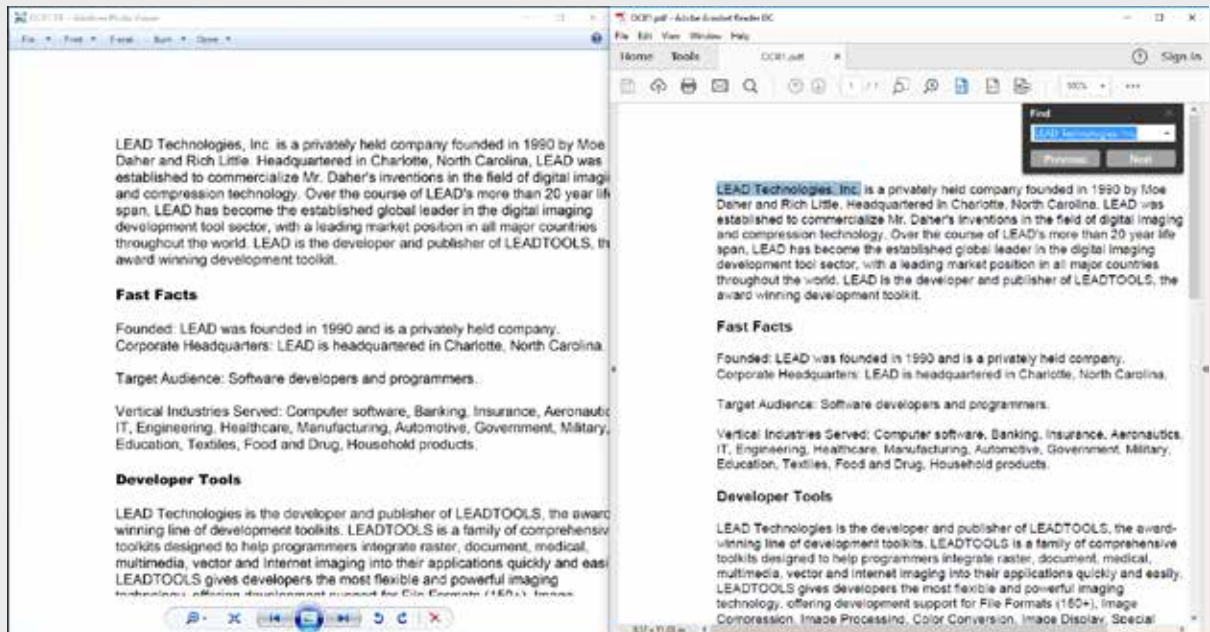
Simple and fast solutions are great, but LEADTOOLS doesn't stop there because we understand that many projects also require customization and more complex tasks. The LEADTOOLS OCR interface is also granular enough to give control over every detail of the process including zones, processing words and characters, even spell checking and modifying the recognition results if necessary. Below, LEADTOOLS is used to recognize the text from a specific rectangle drawn on the image by the user:
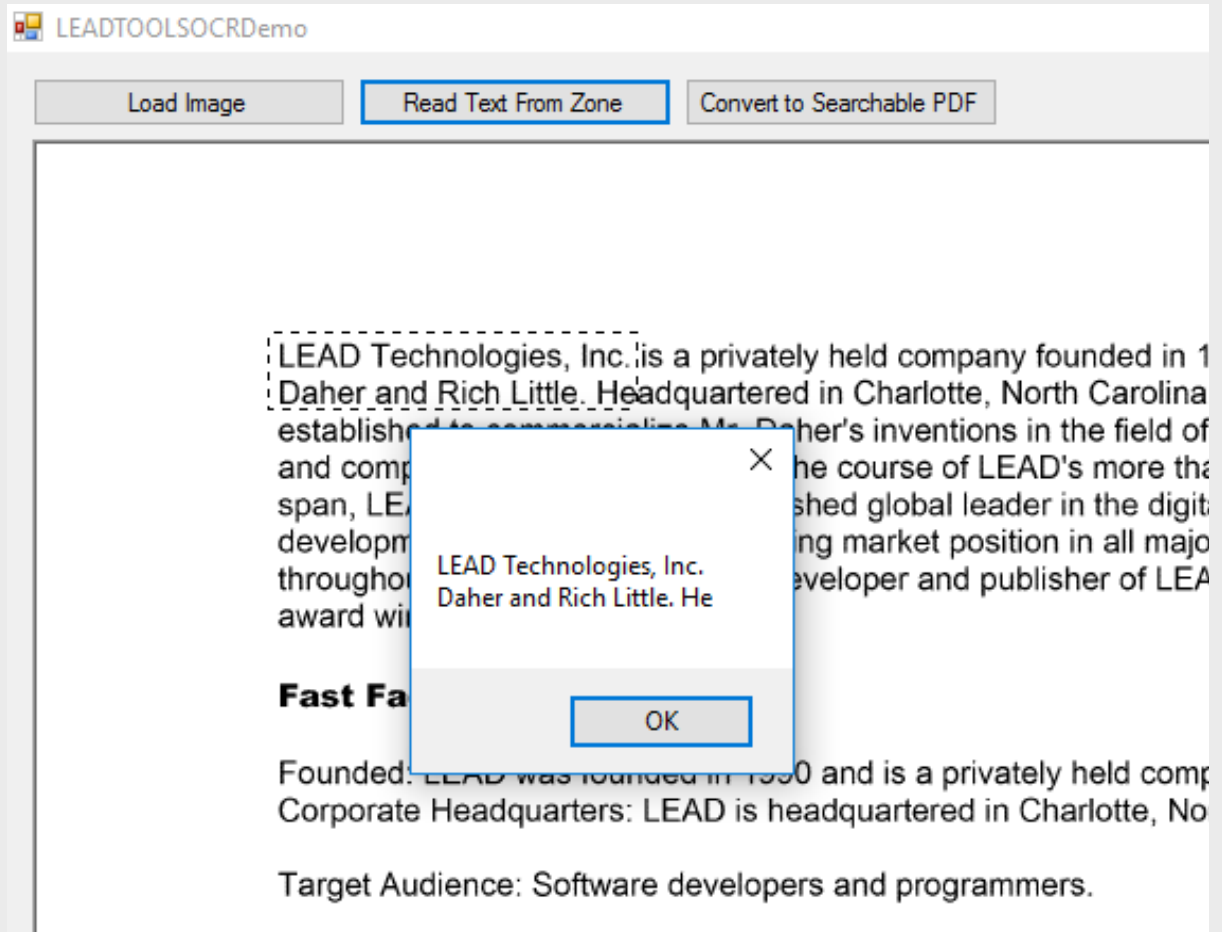
```csharp
// make sure the region isn't empty or the size of the entire image
if (!rasterImageViewer1.Image.HasRegion)
{
    MessageBox.Show("Select a zone in the viewer using the mouse.");
    return;
}

// Create OCR Engine
using (IOcrEngine ocrEngine = OcrEngineManager.CreateEngine(
        OcrEngineType.Advantage, false))
{
    // Start the engine using default parameters
    ocrEngine.Startup(null, null, null, null);

    // Create OCR Document
    using (IOcrDocument ocrDocument = ocrEngine.DocumentManager.CreateDocument())
    {
        // Add image from the viewer as a page in this document
        IOcrPage ocrPage = ocrDocument.Pages.AddPage(rasterImageViewer1.Image, null);

        // Create a zone for the selected region
        OcrZone ocrZone = new OcrZone();
        ocrZone.Bounds = new LogicalRectangle(
            rasterImageViewer1.Image.GetRegionBounds(null));
        ocrZone.ZoneType = OcrZoneType.Text;
        ocrPage.Zones.Add(ocrZone);

        // OCR the image and display text in a MessageBox
        ocrPage.Recognize(null);
        MessageBox.Show(ocrPage.GetText(0));
    }
}
```

LEADTOOLSOCRDemo

| Load Image | Read Text From Zone | Convert to Searchable PDF |

LEAD Technologies, Inc. is a privately held company founded in 1
Daher and Rich Little. Headquartered in Charlotte, North Carolina
established to commercialize Mr. Daher's inventions in the field of
and comp ✕ he course of LEAD's more tha
span, LE shed global leader in the digit
developm LEAD Technologies, Inc. ing market position in all majo
throughou Daher and Rich Little. He veloper and publisher of LEA
award wi

Fast Fa OK

Founded: LEAD was founded in 1990 and is a privately held comp
Corporate Headquarters: LEAD is headquartered in Charlotte, No

Target Audience: Software developers and programmers.

Conclusion

LEADTOOLS provides developers with access to the world's best performing and most stable imaging libraries in an easy-to-use, high-level programming interface enabling rapid development of business-critical applications.

OCR is only one of the many technologies LEADTOOLS has to offer. For more information on our other products, be sure to visit our home page, download a free, fully functioning evaluation SDK, and take advantage of our free technical support during your evaluation.

Sales: +1(704) 332-5532
sales@leadtools.com

Support: +1(704) 372-9681
support@leadtools.com

LEAD Technologies, Inc.
1927 South Tryon Street
Suite 200
Charlotte, NC 28203

About LEAD Technologies

With a rich history of over 26 years, LEAD has established itself as the world's leading provider of software development toolkits for document, medical, multimedia, raster and vector imaging. LEAD's flagship product, LEADTOOLS, holds the top position in every major country throughout the world and boasts a healthy, diverse customer base and strong list of corporate partners including some of the largest and most influential organizations from around the globe.

LEADTOOLS®
THE WORLD LEADER IN IMAGING SDKs

LEAD TECHNOLOGIES
INCORPORATED

www.leadtools.com