CASE STUDIES

Santiago Case Study

Click2Scan Ltd

Using Aspose.Total.Net to convert documents

Martin Peirce, Director, 27th July 2015

About C2S

C2S provide document processing solution for accounts payable and digital mailroom.

At the core of our solution are two systems - ABBYY OCR & FileDirector Document Management. Both systems process and handle documents that are images like PDF or Tiff perfectly. ABBYY OCR cannot recognise existing text document like Excel or Word and FileDirector will only allow annotations of scanned documents like Tiff.

Problem

C2S have always struggled to find a consistent solution for document handling that can carefully and accurately convert documents. We then came across the Aspose.Total for .NET.

Solution

Having previously used PDF# and other open source libraries, we were more often than not left disappointed by the performance. Our clients require bullet proof processing and accuracy.

During the development of our Santiago tool, we had a specific customer requirement that made us search and find a perfect solution. Santiago is a scheduler that monitors email accounts and downloads attachments that meet set rules. The attachments are then forwarded into a workflow. Santiago effectively feeds a digital mailroom or accounts payable solution.

The customer in question received attachments from over 500 branches around London. The attachments are agendas and minutes of meetings and can be in any format but are usually typed in Word or Excel. We needed to forward these into our FileDirector solution, fully indexed into the correct queue but the customer asked to be able to annotate the documents – add text, arrows, highlights etc. This would only be possible if we converted the documents into Tiff.

We have now built the Aspose.Total.Net library into Santiago with excellent results.

Experience

The Santiago email downloader solution has been designed to be completely modular and therefore extremely flexible. The system is designed to download emails, extract information and save it as required. It is capable of saving not just the attachments of an email, but also the body (both html and plain text) and the metadata (sender address, recipient addresses, subject, date, etc) and saving this in a range of formats.

Our primary requirement that led us to Aspose.Total for .NET was the need to convert documents (mainly MS Word and Excel formats) into TIFF. We experimented with several different libraries but Aspose.Total for .NET was by far the best solution.

Santiago operates by downloading emails as .eml files into a user designated folder, from this folder the Data Extraction module picks up the .eml file, extracts the required documents and saves this to a further designated folder. Finally, the File2Tiff module picks up the extracted documents, performs the conversion and again saves them to the final location where they can be collected by the FileDirector system for archiving. The workflow for this process is shown in Fig. 1.
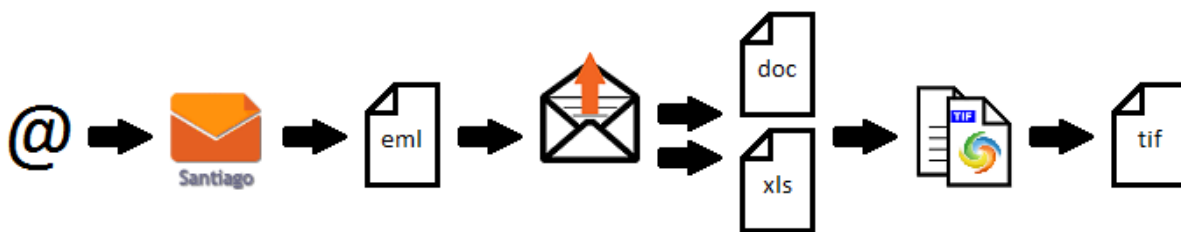

Fig. 1. File2Tiff workflow

The File2Tiff module examines the extension of the file to be converted and selects the required conversion system:

```
fileType = Path.GetExtension(filePath).Replace(".", "").ToLower();

if (FileTypeExtensions.File2TiffDoc_Extensions.Contains(fileType))
{
        ...
```

<<Product>> Case Study

```
        }
        else if (FileTypeExtensions.File2TiffXls_Extensions.Contains(fileType))
        {
                ...
        }
```

File2Tiff is then able to utilise the Aspose.Words and Aspose.Cells libraries to perform the conversion, making use of the Document and Workbook classes respectively, using only a few lines of code:

Aspose.Words

```
        Document doc = new Document(filePath);
        ...
        doc.Save(Path.Combine(OutputPath, outputFileName), options);
```

Aspose.Cells

```
Workbook book = new Workbook(filePath);
Worksheet sheet;
SheetRender sheetRender;
Bitmap bitmap;

...

for (int i = 0; i < book.Worksheets.Count; i++)
{
        sheet = book.Worksheets[i];
        sheetRender = new SheetRender(sheet, imgOptions);

        for (int j = 0; j < sheetRender.PageCount; j++)
        {
                bitmap = sheetRender.ToImage(j);

                if (bitmap != null)
                {
                        bitmap.Save(Path.Combine(OutputPath,
                        string.Format("{0}_{1} - Page {2}",
                        outputFileName, i + 1, j + 1)));
                }
        }
}
```

A particularly useful feature of both libraries is the ability to specify image saving options. This allows us to provide the end user with the flexibility to specify the resolution at which they wish to save the document, selecting between saving in colour or greyscale and choosing particular pages of a document to convert:

```
Aspose.Words.Saving.ImageSaveOptions options = new
        Aspose.Words.Saving.ImageSaveOptions(Aspose.Words.SaveFormat.Tiff);
options.PageIndex = UserSelectedFirstPage;
options.PageCount = UserSelectedPageCount;
options.Resolution = UserSelectedResolution;
```

<<Product>> Case Study

```
options.ColorMode = UserSelectedColourMode;
```

Summary

The Aspose.Total for .NET package offers significantly more functionality than we have implemented thus far. We will continue to enhance the Santiago system further to harness the power of this product, as well as re-visiting the existing modules to ascertain whether they could benefit from the addition of Aspose.

<<Product>> Case Study