vision docs

1 Mint VisionDocs

1.1 Introduction

Mint VisionDocs is a service that extracts information from documents, such as PDF's, and makes it searchable and available to other applications.

The software can recognise standardized forms, extract relevant information, and publish it to external systems – or just allow you to search through the information.

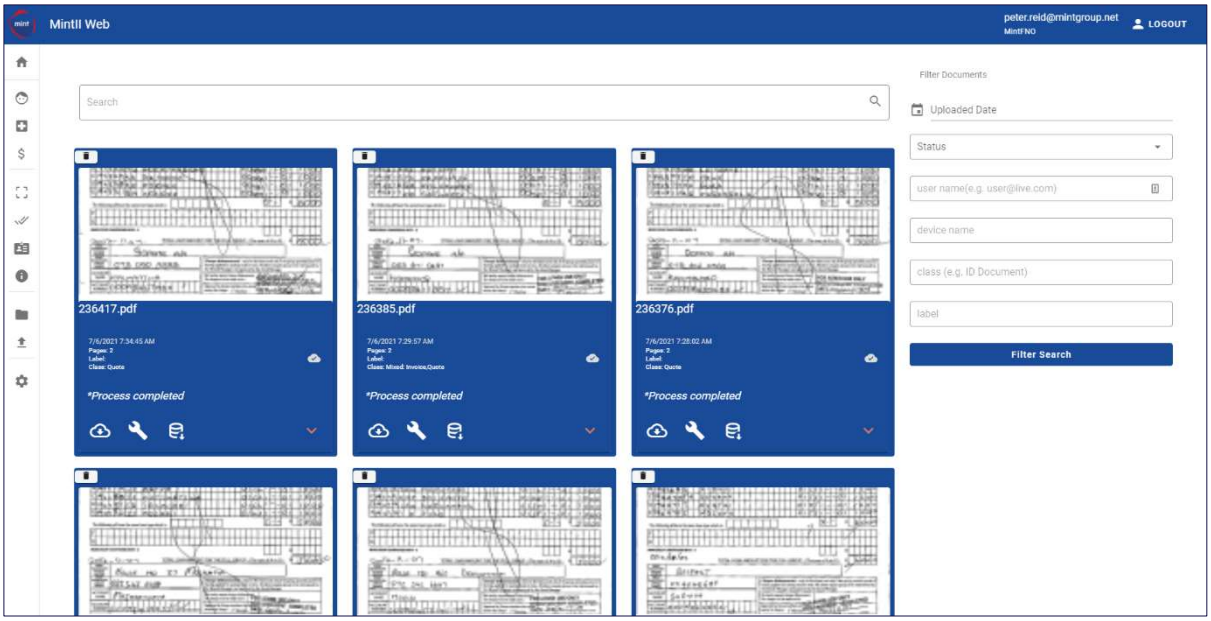

Figure 1: Documents in the system

The system can understand and analyze pdf documents and images, and is capable of extracting printed text, handwriting and image content from the documents.

1.2 Ingestion

The system is optimized for millions of documents to be submitted into the platform. Documents can be inputted into the system in a multitude of ways.

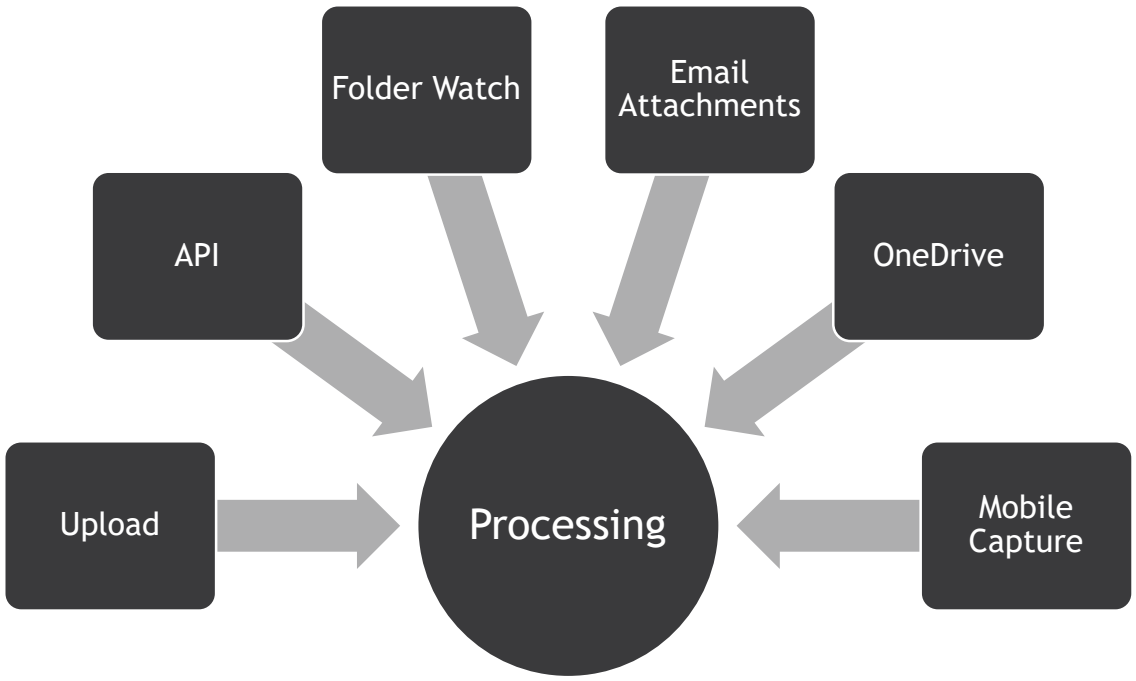

Figure 2: Multiple ways to ingest

1.2.1 Upload

Documents can be uploaded directly into the system using the Web Front-End. For smaller volumes of documents, this is the easiest way to test and use the system. This also provides an interface into the rest of the system.

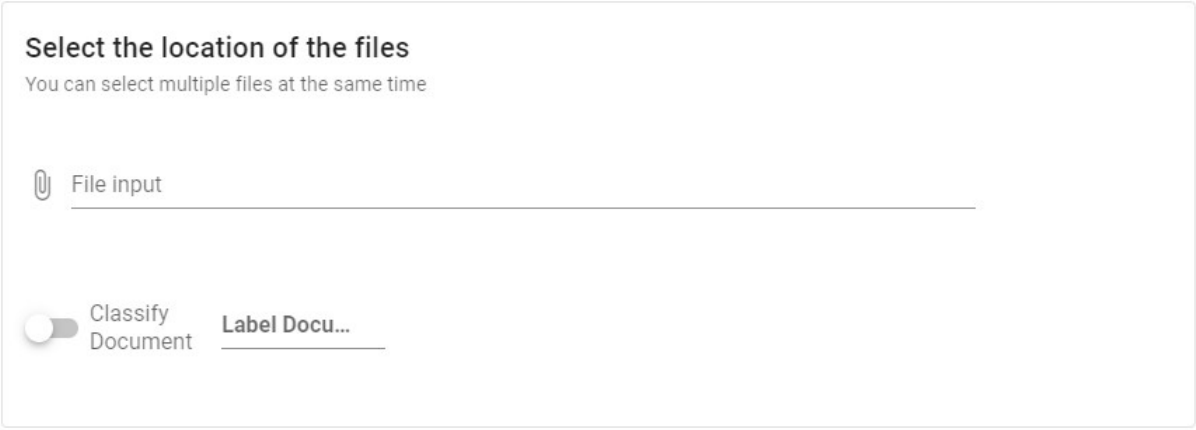

Figure 3: Uploading Documents

1.2.2 API

The system has a full REST API, which allows integration into any system. The API allows for full processing of documents including Uploading, Querying, and all other operations on the documents.

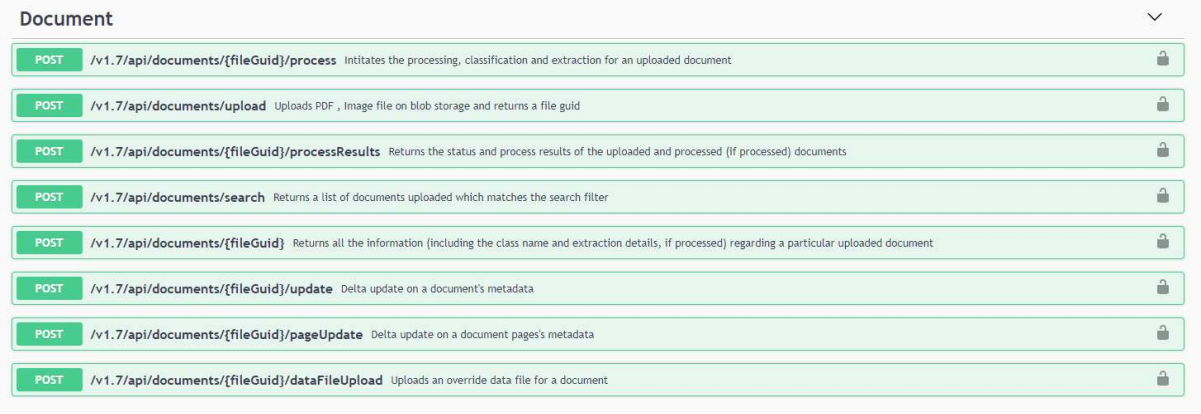

Figure 4: REST API for document control

1.2.3 Folder Watch

The Mint Vision Application, downloadable from the Microsoft Store, can watch a folder location for new documents – and upload and process them as they arrive in that location.

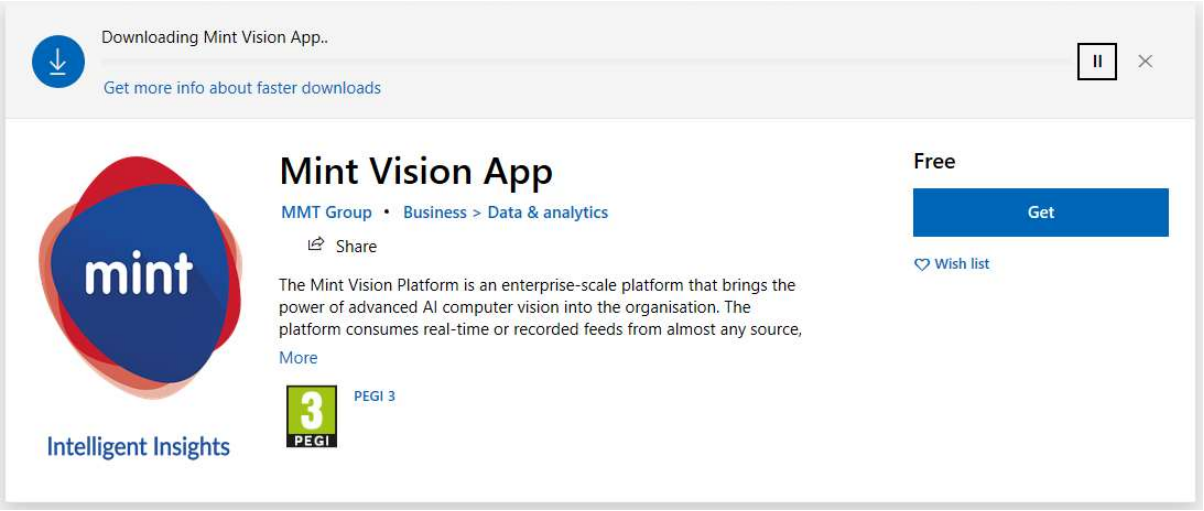

Figure 5: Mint Vision Application in the Microsoft Store

The application can be run on multiple stations, and enhanced metadata can be set per document or path.

1.2.4 Email Attachments

The system can auto-create an email address that can be watched for any incoming emails with attachments and, if they come in, will strip out the attachment and submit it for processing

1.2.5 OneDrive

A special folder can be created in OneDrive and monitored for any new documents or changes. When new documents arrive, they are picked up automatically and moved into the system.

1.2.6 Mobile Capture

The VisionAutomate system can be used to ask customers to submit documents from their mobile, by uploading directly from their mobile or by taking a picture of a physical document with their camera phone.

1.3 Processing

Once a document or image is in the system, it begins to be processed. Processing undergoes the following sequence of steps.

Figure 6: Document Processing Pipeline

Each step is described in detail below.

1.3.1 Page Splitting

Once the document is ingested into the system, it is split into its component pages. Each page is then sent for further processing along the pipeline. Pages can be viewed, downloaded as images, or viewed as thumbnails in other systems.

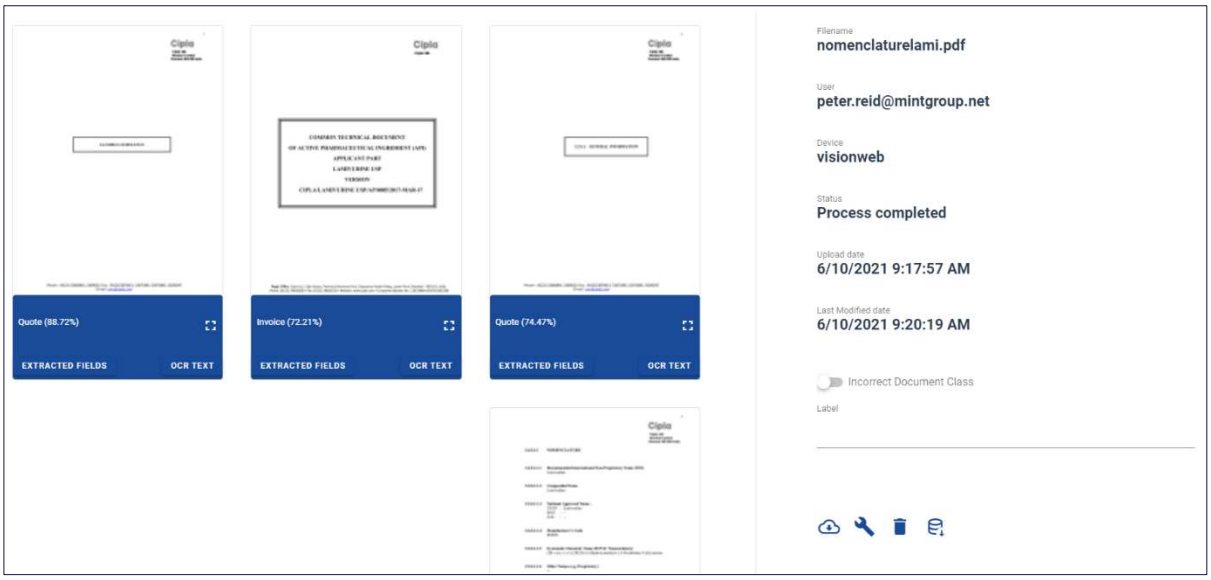

Figure 7: Pages split, and thumbnails extracted

Through the API, each page can be accessed, downloaded and searched as a single entity, as well as the document as a whole.

1.3.2 OCR

OCR (Optical Character Recognition) is then performed over each of the pages. The words and characters are lifted from each of the documents and stored alongside each page and each document. This text is used in subsequent pages for further processing.

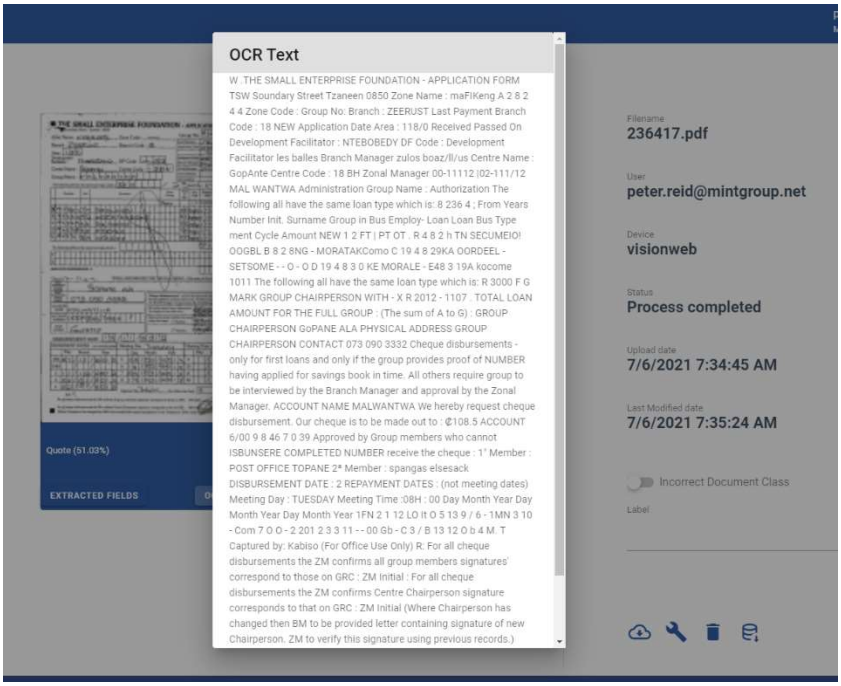

Figure 8: Extracted Text for a document

Text that is extracted from the pages is then added to Microsoft Azure Cognitive Search – a search engine that can search millions of pages in real time. This allows all pages that have been processed to immediately be searchable from the web front-and, and through the API.

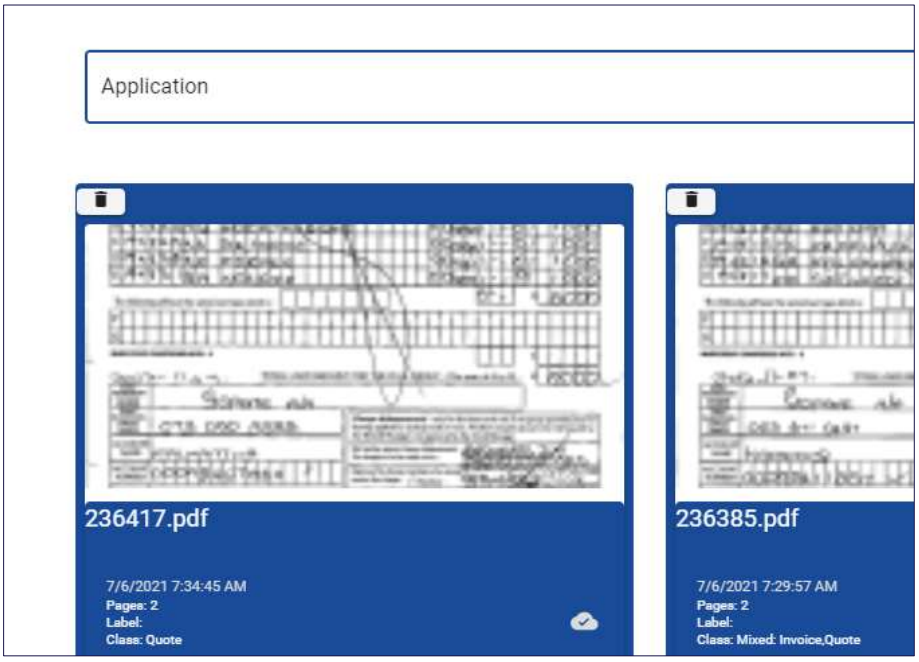

Figure 9: Instant Search across documents

1.3.3 Classification

If the tenant has been configured and trained for it, then the pages are classified – recognised using advanced artificial intelligence, against a pre-trained model.

With a starting set of only 20 samples, we will train a classifier to recognize text, logos, and other marks on the document (no matter how diverse and distributed) and decide what type of page/document it is.

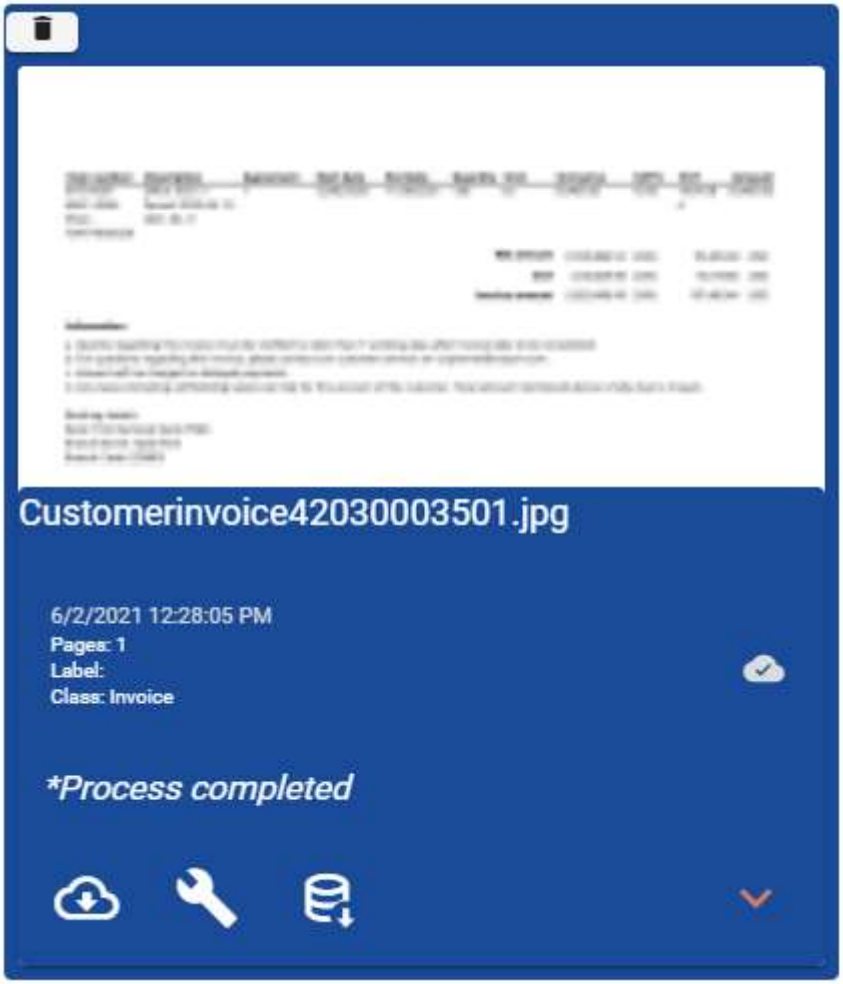

Figure 10: This document has been classified as an Invoice

1.3.4 Extraction

If the page is of a known type, and an Extractor has been trained on that type, then known fields are automatically extracted from the document. This uses another type of advanced AI machine to figure out which properties are present on the document, regardless of whether they are present in the same place on the document, or which form they take. The AI network uses linguistic and visual analysis, as well as other clues, from trained document samples to extract the correct fields and their values into key-value pairs.

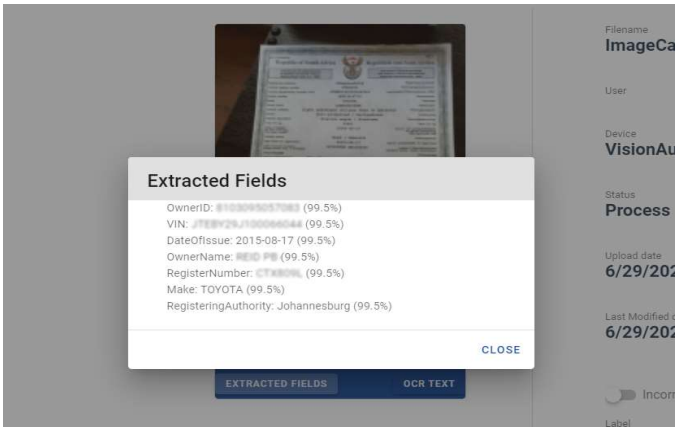

Figure 11: Extracted Fields from a Known Document

1.3.5 Post-Processing

Documents can have custom post-processing steps inserted, to perform special operations on them. This allows for any type of "Special" documents to have additional processing performed on them, once they are through the system. For example, once Invoices are processed the line-items for invoices are extracted and parsed into a CSV file, ready to be used in an ERP system or balanced to a finance solution.

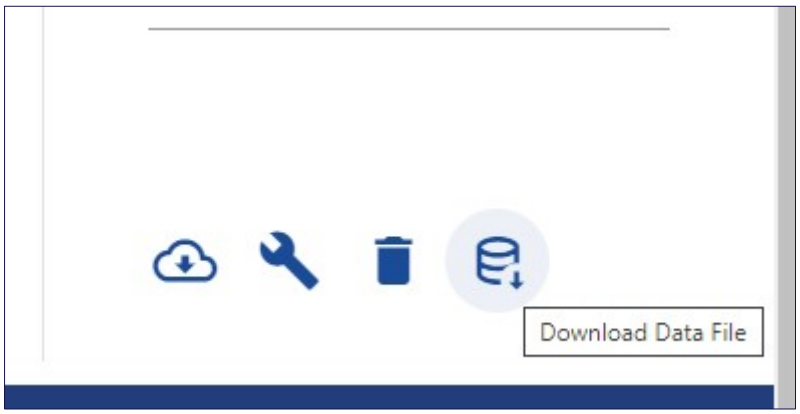

Figure 12: Post-processing can produce a data file

1.4 Additional Features

The Mint VisionDocs system has a host of additional features that can be used as part of a subscription. This document lists some of those features.

This is not an exhaustive list, and customers are encouraged to reach out to us for more information or a demo of the finer points of the product.

1.4.1 Integration

The platform includes several points of integration, for inserting documents into the system, retrieving them or automated pushing of documents into systems such as SharePoint or Teams. Built on the Mint Vision Platform, the system has a full-functioning REST API for developers to integrate with. In addition to this, all actions performed within the platform raise Events on Azure Event Hub, which can be listened to anywhere in the world and responded to.

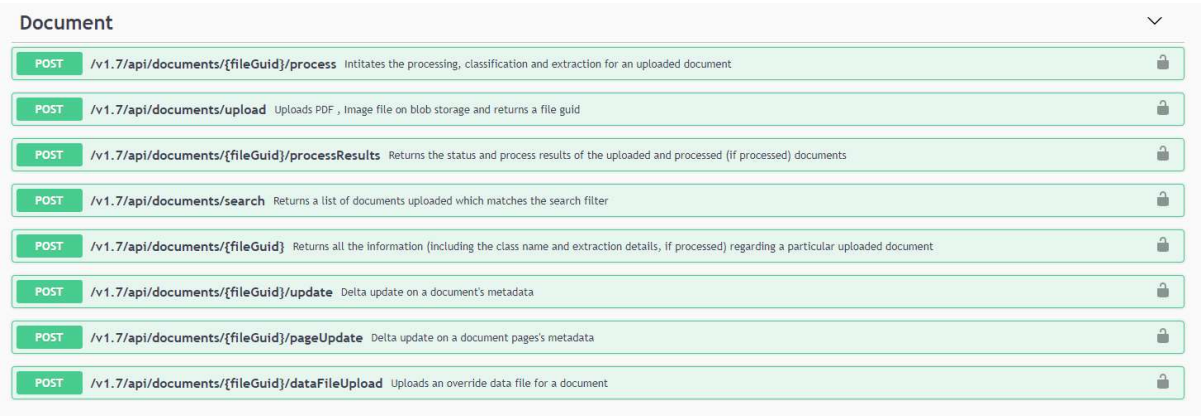

Figure 13: REST API for Document Control

As a demonstrator, Mint has written a basic event listener that publishes processed documents such as Invoices to a Teams Channel. Note that this is just an example project to demonstrate the power of Events being raised at every point in the document processing pipeline.

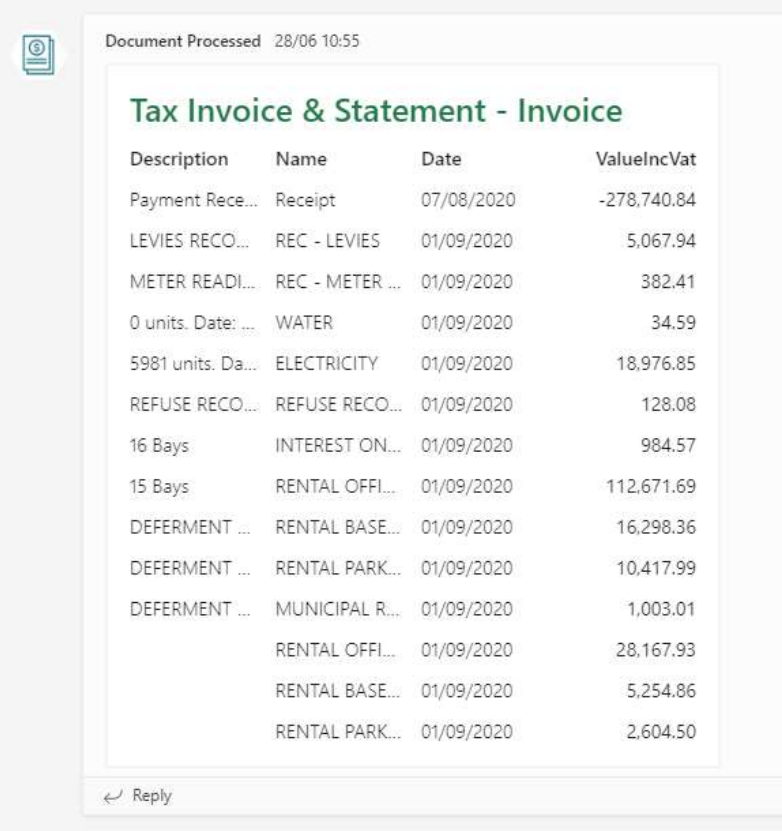

Figure 14: Example listener, which inserts processed Invoice Items into a Teams Channel

To assist developers with integration, comprehensive documentation and example code is available for them to download, copy, and learn from. The Sample Projects and code cover all of the commonly requested and implemented features and functionality to get developers started on our easy integration platform.

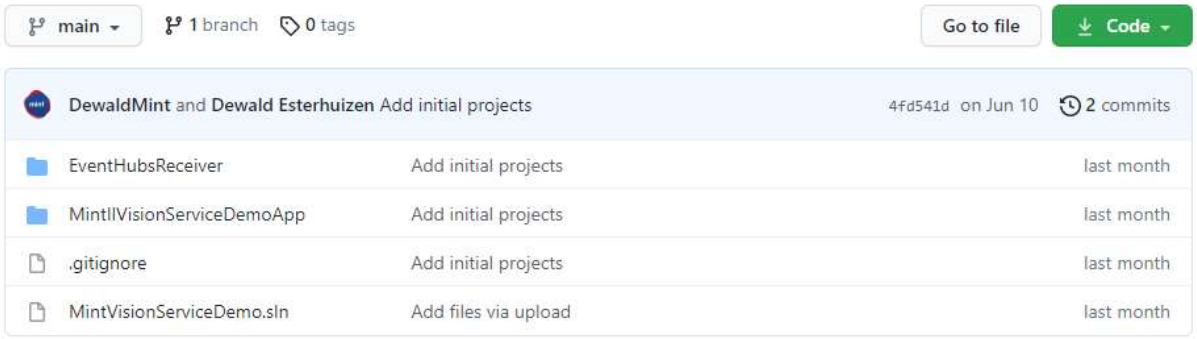

Figure 15: Sample code for developers

1.4.2 Continuous Improvement

The VisionDocs web interface contains tools and techniques to allow the system to learn and improve over time. Specifically, the classification of pages and the information extracted from

Known Pages, can be indicated as incorrect. This then moves back into our team of specialists who review, and change/add the samples so that the system improves over time.

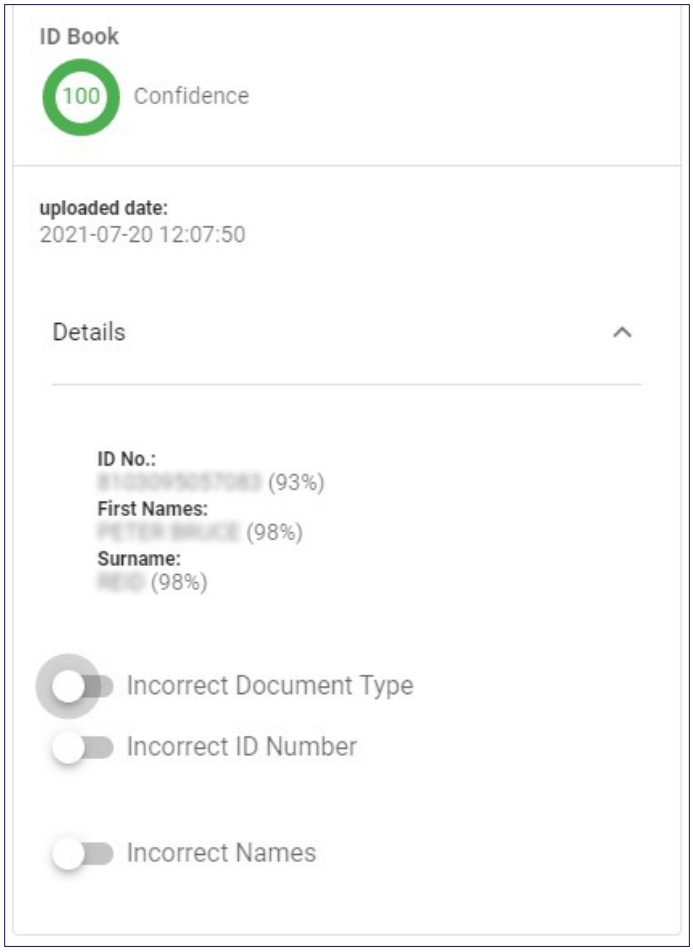

Figure 16: Indicate which fields are incorrect for further training

1.4.3 Known document processing

Certain document types are already trained into the system, and produce more results than standard OCR and classification results. Invoices, for example, provide invoice line items and extracted fields such as total and VAT number.

ID Documents can include face comparisons for likeliness metrics, and even a call into the Government Database to verify that they user is who they say they are.

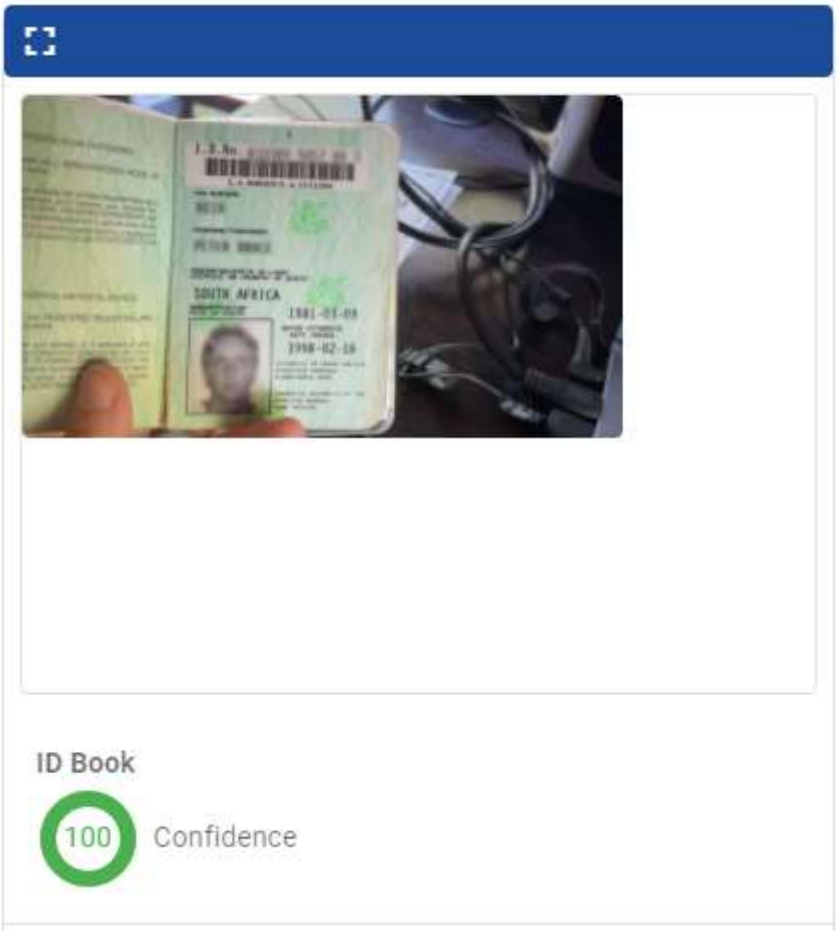

Figure 17: ID Documents are also submitted to the Government Database

1.4.4 Advanced Dashboards

Although the product has standard dashboards, available for every user, we can customise new dashboards for your specific requirements to show you the information you need, when you need it.

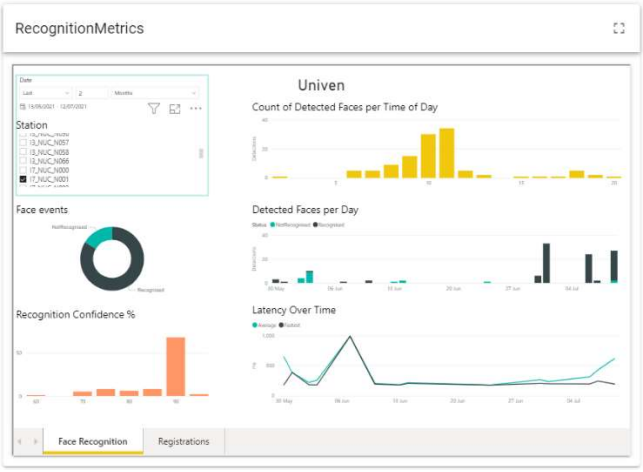

Figure 18: Sample Dashboard

2 More Information

For more information, visit our dedicated Kiosk Page on our website at
https://www.visionaisuite.net/products/visiondocs-ocr-document-scanner/

Figure 19: VisionDocs Website

Or contact visionsupport@mintgroup.net