



Parascript® Document Classification

Parascript Document Classification rapidly identifies and organizes document-based information to support compliance, discovery, and data management applications. By providing an easy-to-use analysis tool built on a powerful production SDK platform, Parascript places document auto-classification in the hands of business users. On an ongoing basis, high volumes of documents are dynamically classified with high accuracy and low error rates without users needing any programming experience.

Parascript Document Classification organizes documents and associated data faster with greater accuracy for increased productivity at a lower cost.

Classification Made Easy

Parascript Document Classification makes identifying and organizing documents a simple and straightforward process without knowing anything about the documents that need to be organized. Once they are classified, you can add simple workflows to add metadata and even use the data within the documents to add more-descriptive searchable data.

Parascript leverages its advanced data location and extraction techniques along with its state-of-the-art image analysis capabilities to develop Document Classification. It automatically applies multiple techniques including algorithms that process visual, text, or glyph-like elements—such as signatures or logos—in order to classify documents based upon all the available information on the page.

Classification employs a variety of classification methods and is based on three different types of classifiers: Content Classifier, Visual Classifier, and Combined Classifier. Content classification employs text-based features of documents. Visual classification uses “feature extraction” based upon key visual characteristics of the documents. Combined classification uses both of them and synthesizes the results for the best answer. All of this allows results to be highly accurate with very low error rates.

Leveraging All Data

Classifies documents based on any data. Unique to Parascript Document Classification is the capability to organize documents not only based on features and text, but also on imagery and handwritten information on the document, including the presence of signatures. Regardless of how the information is presented or what it contains, classification is based upon all of the available information on the document, not just a select subset of that data. Documents that have handwritten social security numbers or other sensitive information can be included in the classification workflow and easily identified. If presence of logos or other imagery can help to separate one document class from another, that can also be leveraged.

Discovery by Auto-classification of Unknown Documents

Many document classification solutions require Subject Matter Experts (SMEs) familiar with the organization's existing "information taxonomy" to assemble samples of targeted documents. This time-consuming preparation from key staff is both resource intensive and costly. Parascript Document Classification processes documents, taking them from an unknown to known state in minutes. After simply importing a volume of documents, Parascript Document Classification automatically groups the documents based upon content or feature likeness, all without any prior preparation. Grouped documents can then be reviewed and used as samples to perform classification. This ensures that every classification project is comprehensive without the extensive overhead of preparation costs.

Simple to Use

Dynamically groups documents — Import any volume of documents without any preparation and the software automatically classifies them. Visually move documents from one group to another and name them to create classes, all without any programming experience. Fine-tuning document classes is as simple as correcting results by moving results from one class to another and then re-running the task so that the software learns the right class for the document.

Create complex classification rules with a visual rule-builder that can be combined with text or visual classification to create special classes of documents. Getting to know your documents and enable improved governance and document search has never been easier or more accessible.

Easy API Integration

Easily integrates with existing and new systems — Once classes have been created, production use of Classification can be easily integrated using the .NET API.

Key Applications

Compliance — Easily and comprehensively scan documents for any type of sensitive information. Once identified, the software can even redact the information.

Discovery — Automate the process of grouping a collection of documents, and then use this information to process an entire volume of documents to support legal or compliance needs.

Migration — Take collections of documents on a file share or third-party system and easily organize, extract, and apply key metadata to simplify and organize documents into a consolidated content management system.

Reduction of ROT - Redundant, Obsolete, and Trivial — Identify duplicate documents and documents that don't need to be preserved and easily remove them.

Key Benefits Summary

With Parascript Document Classification 2.0, Parascript advances classification far beyond all batch classification approaches, which are limited to organizing pre-defined sets of documents in an established order. Unlike other auto-classification solutions, Parascript Document Classification provides the following unique benefits:

- **Offers high accuracy** classification with low error rate
- **Applies multiple document classification techniques automatically** including algorithms that process visual, text, or glyph-like elements
- **Groups documents automatically without samples**
- **Employs trainable machine-learning** so that with a few samples, the software automatically learns the features of a particular document class
- **Provides accessible and understandable UI for business users**

Technical Product Specifications

Included Software Components

- Document Definition Studio

Requirements

- **Document Definition Studio:** Windows 7 or 8
- **Server:** Windows Server 2012 or 2014
- **OS:** Use 64-bit OS for high performance systems to support more memory

- **Disk Space:** Hard Disk Space 1GB of free disk space

Input

- **Image Format:** Black-and-white or grayscale and includes all of the following formats: TIFF, JPEG, PNG and PDF.
- **Image Resolution:** 200-300 DPI

Output

- XML file with document file name and class assignment

- Document with class as part of file name
- MS SharePoint with file and class assignment as metadata
- RDBMS with document as blob and class as record column
- Other output options available

Availability

- Parascript Document Classification is available standalone or as an optional module within *FormXtra*.