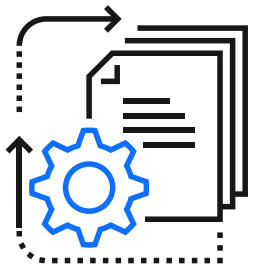
Solution brief

IBM Business Automation Content Analyzer

What it is, what it does, why it matters



IBM Content Analyzer in 10 seconds

IBM Content Analyzer is an intelligent Capture-as-a-Service (CaaS) system that allows you to:

- Understand data in structured and unstructured documents the way a person would, using AI
- Lessen the dependency on IT-created templates so that you are empowered to make rapid changes
- Get more out of your existing data capture, content management or process systems
- Scale with the flexibility of a cloud-based API

Why Intelligent Capture-as-a-Service?

Humans are able to look at a document and easily classify information using context clues within the document. Software that's used to capture information in documents doesn't always have the ability to contextualize this data and logically organize it. Intelligent capture solutions use AI to classify and convert unstructured content into structured and actionable information much like a human, but faster.

CaaS simply refers to a capture software solution with a Software-as-a-Service (SaaS) hosting and purchasing model, so you can start small with a low cost of entry and scale as needed. Software is hosted and managed on a vendor's cloud, which minimizes IT involvement and improves time to market.

Intelligent Capture-as-a-Service allows you to take a layered approach to data capture. With APIs built on microservices architecture, you can easily :

- Embed capture services into existing applications
- Add more value to your current technology stack, including document capture, content repositories, information gover nance, task automation, workflow and decisioning.
- Extract to a data lake for reporting, visualization, analytics and machine learning with tools like IBM Watson.

53%

of Chief Data Officers believe lack of data access is the biggest barrier to success for their companies.*

How IBM Content Analyzer can help

Intelligent CaaS solutions, like IBM Content Analyzer, give you the flexibility, ease of use and classification and extraction capabilities you need to optimize your digitization and data analytics strategy. You're able to drive immediate business value through a cloud-based service that applies AI to enable big data analytics and streamline complex business processes.

IBM Content Analyzer uses AI to classify and extract unstructured data from documents. It can be embedded into existing content or process systems to extend the value of your current systems. The user interface is intuitive for a business analyst who understands the documents. For IT, the API is modern and easy to use.

Key features and benefits:

- Document classification configurable by a business analyst. Unlike other data classification systems, no scripting or templating is required. You can run testsand modify data classification on tology with no IT or engineering resources needed.
- With our AI algorithms, you are ready in minutes, not weeks or months. Speed up your time to market for the most value.
- Classify and extract data from highly variable content. You can't control the format of all the documents that come from your customers and partners. Get data from many variations of file types such as PDF, TIF, PNG, and JPEG with minimal input required.
- Use as a standalone service or extend the value of your automation platform. You can use Content Analyzer as an API directly in your applications or as part IBM's digital business automation platform.
- Understands key-value pairs in documents. After simple training, Content Analyzer reads your documents, then it uncovers and returns semantically normalized data.

Content Analyzer can be used in most use cases, but is particularly good when you need to:

- Quickly identify key information in contracts and documents for compliance
- Digitize processes that rely on exchanging documents such as account opening, insurance claims or payment processing to reduce labor cost, accelerate turn-around time and improve customer experience
- Mine data from existing docu ments to a data lake for reporting, visualization, analytics and machine learning with tools like IBM Watson.

What makes IBM Content Analyzer unique

Top reasons you should consider Content Analyzer as part of your data capture strategy

You need to ...	**Automate business processes with RPA and BPM.** Your process relies on the exchange of documents and the required data is locked within a document that contains essential data.	**Identify sensitive data for regulatory compliance.** Your organization is unable to automatically identify content that may contain protected or sensitive data.	**Understand and evaluate contracts and other types of documents using AI.** You have contracts that resist analysis because they are stored as images and have highly variable layouts.	**Setup data capture faster with less effort and maintenance** Managing templates and large training sets can be time and resource intensive, meaning less agility for your organization.	**Process many documents without backlogs.** You can't keep up with heavy document workloads due to limited processing resources in your data center.
IBM Content Analyzer can ...	**Automatically extract data from documents** so that your process has the necessary data with less human intervention. **Integrate into any automation or AI tool,** allowing you to effectively integrate your current systems with AI and Automation technologies.	**Automatically identify the document type and flag sensitive data** so that downstream systems can ensure regulatory compliance.	**Process thousands of documents,** including data from both structured and unstructured content, using AI and semantic normalization. It can also call Watson services to handle additional processing.	**Be configured by a Business Analyst without the need to involve a technical expert.** No templates are needed, and flexible algorithms can accommodate many different formats. This means that configuration takes days instead of weeks or months.	**Offers cloud-scale resources.** Because it's hosted on the IBM cloud, it can handle heavy workloads and rapid response without requiring resources in your datacenter.

How IBM Content Analyzers works

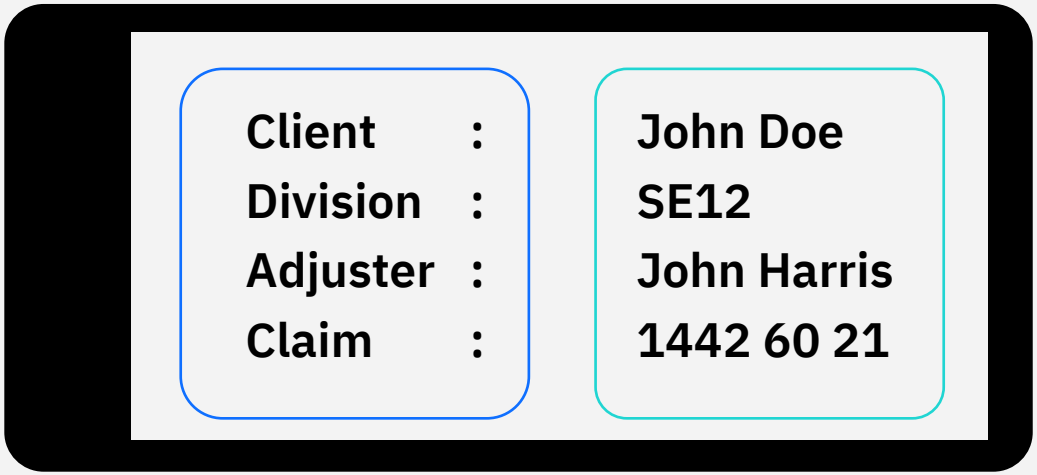
IBM Content Analyzer helps you to optimize your data capture by:

First, making it easy for your business users to configure, process and test document data

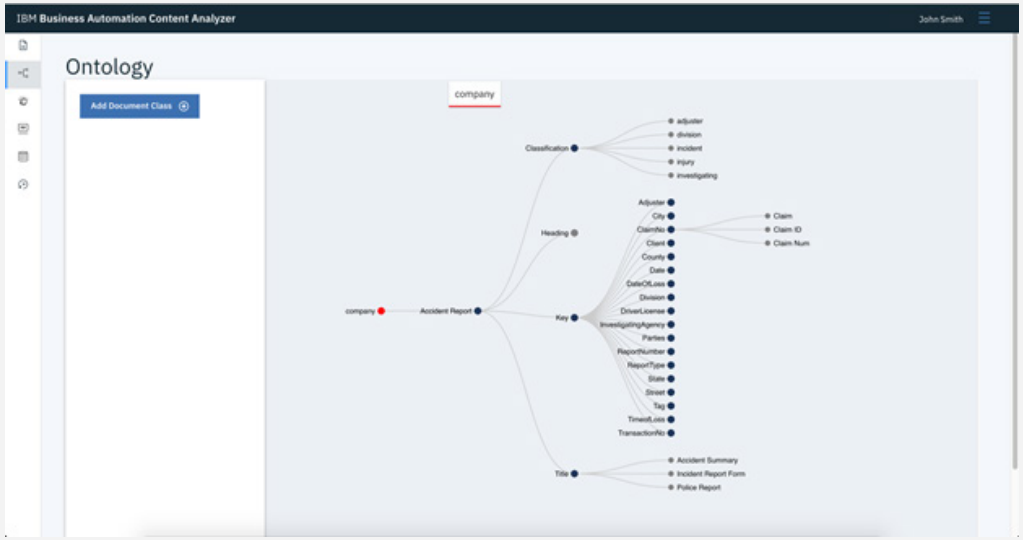
IBM Content Analyzer allows you to extract and classify document data through the use of an intuitive ontology framework. An ontology allows you to classify documents and identify key-value pairs (KVPs) without the need to develop document templates, which requires technical expertise or frequent maintenance. This means that a business analyst – the one who understands the documents – can configure the ontology, upload documents for testing output and modify the ontology to achieve optimal data results, all without the need to create many templates or involve a technical expert.

What is a KVP?

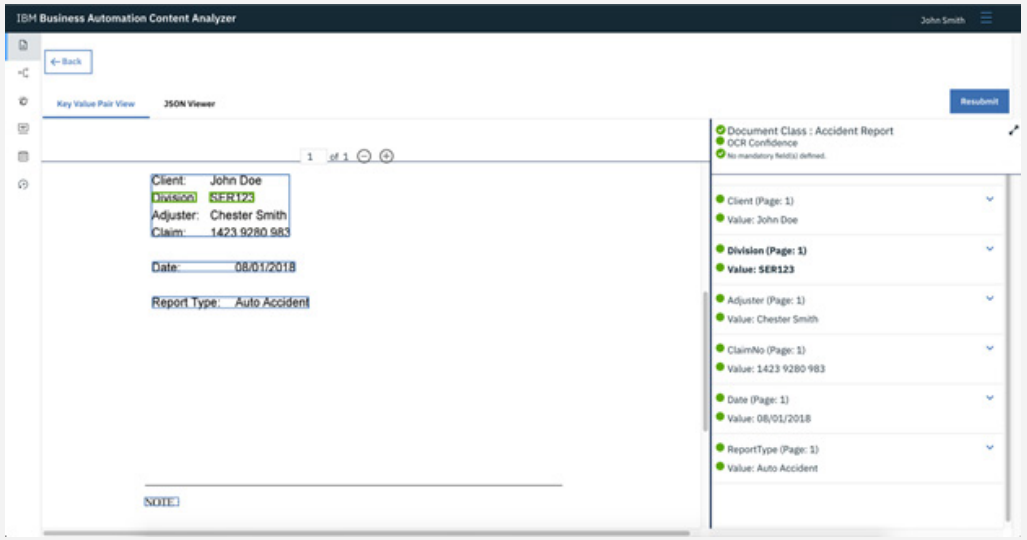
Simply, a group of two data items that should be grouped together. For example, the circle to the right contains keys, and the circle to the left values. Client/John Doe is a key value pair.



A business analyst accesses the ontology builder and creates a list of document types, the data fields and initial clues to teach the system.



After uploading a sample document, the document is classified and the data is identified. The user can correct errors to teach the system and improve results.



Second, making it easy to deploy and connect quickly – minimizing time to use

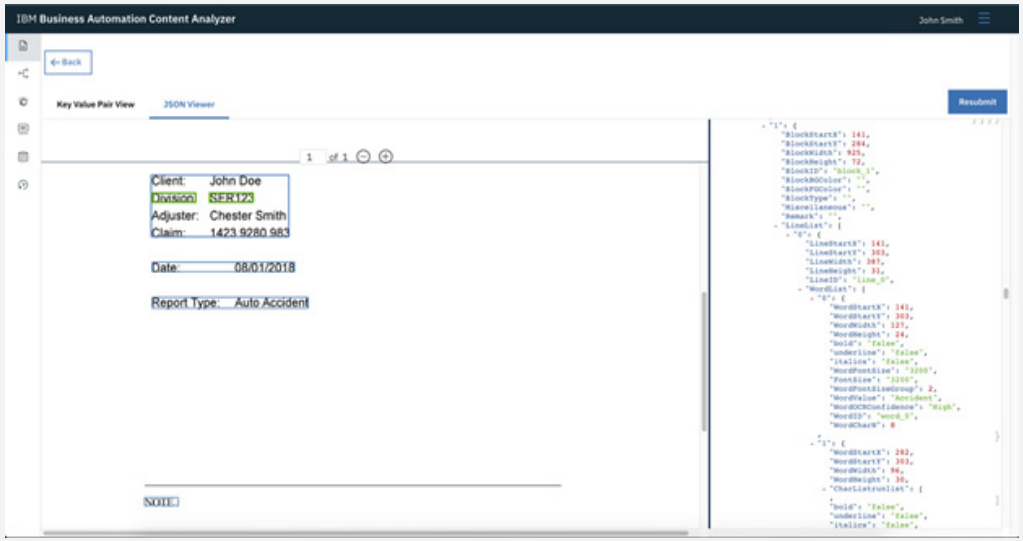
IBM Content Analyzer is a REST API built on microservices and designed in the cloud. This means that development timeframe to embed into your system is drastically reduced, especially if embedding into other cloud-based applications. Your IT group, or IBM Services, adds Content Analyzer into your application.

A business analyst accesses the ontology builder and creates a list of document types, the data fields and initial clues to teach the system.

After uploading a sample document, the document is classified and the data is identified. The user can correct errors to teach the system and improve results.

Finally, making it easy to export and apply data– wherever you need it

Data is formatted into a consistent JSON file, regardless of document type. The data is returned when a document is processed along with a searchable PDF file. This allows you to empower your applications and automation platform with full and complete data to improve your business function.



Example showing JSON output that is sent via API call to your application

To see how it works:

Watch this
4-minute demo

Take the 10-minute
guided tour

Need to talk to an expert?

Schedule a 30-minute
consultation

Sources:
https://www.datascience.com/blog/2018-data-science-predictions

© Copyright IBM Corporation 2018

IBM Corporation
New Orchard Road
Armonk, NY 10504

Produced in the United States of America
June 2018

IBM, the IBM logo, ibm.com, and IBM Watson are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

All client examples cited or described are presented as illustrations of the manner in which some clients have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual client confgurations and conditions. Contact IBM to see what we can do for you.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.