

Automate document processing with AI

EXTRACT STRUCTURED DATA FROM
DOCUMENTS WITH SPEED, FLEXIBILITY,
AND ACCURACY USING AMAZON Textract

Use machine learning to instantly “read” your documents



Documents are important—for records, communication, collaboration, and even transactions. Yet millions of times per day, users rely on information that is locked in documents.

This is a broad challenge across industries—affecting anyone who sends and receives paper documents, faxes, digital files, and/or needs to keep records. It especially affects highly regulated industries including finance, healthcare, and government. Processing documents requires manual effort, traditional optical character recognition (OCR) software, and rules/template-based extraction. But this is expensive, error-prone, time-consuming,

and voluminous. Take, for example, the 16.3M US mortgage applications from 2016¹ and the nearly quarter of a billion W2 tax forms expected to be processed in the US in 2018.²

Amazon Textract uses artificial intelligence to “read” documents as a person would, to extract not only text but also tables, forms, and other structured data without configuration, training, or custom code.

In this eBook, learn how Amazon Textract can help you extract data with accuracy, flexibility, and speed.

1. Mortgage Bankers Association 2016 HMDA
2. IRS <https://www.irs.gov/individuals/w-2-verification-code>





What is Amazon Textract?

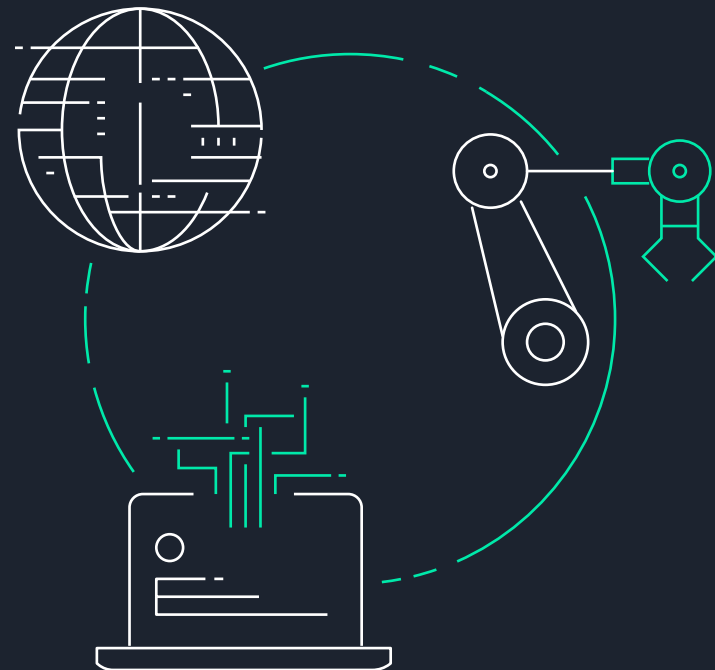
Amazon Textract is at the heart of solving an old problem: decreasing the manual effort needed to extract and process information from documents. Using machine learning (ML), Amazon Textract automatically “reads” virtually any type of document to accurately extract text, forms, and tables—without the need for any manual effort, ML expertise, or custom code.

Amazon Textract goes beyond simple Optical Character Recognition (OCR) to identify the contents of fields in forms (such as key-value pairs), information stored in tables, and the context in which the information is presented, such as a name or social security number from a tax form.

The extracted text and data can then be easily used to build smart searches on large archives of documents or to automate downstream business processes. It can be loaded into a database for use by applications, such as accounting, auditing, processing, and compliance software, or as part of Robotic Process Automation (RPA).

Amazon Textract is:

- **Fast:** Works automatically and eliminates the need to write code for data extraction—meaning you can extract the details you need quickly.
- **Flexible:** Extracts text and data from virtually any type of document—including forms and tables.
- **Accurate:** Uses machine learning to extract data—reducing manual human error.



What you can do with Amazon Textract



Create smart search indexes—simplify workflows and tasks like searching for names or policy numbers in a sea of documents



Build automated document processing workflows—process millions of document pages in hours and send key pieces of information to the relevant system or team



Analyze trends in your data over time using additional ML techniques



Perform RPA functions—to automate workflows and improve business agility

Amazon Textract features and capabilities:

- Optical Character Recognition (OCR)
- Form extraction
- Table extraction
- Bounding boxes
- Adjustable confidence thresholds

Next, read about how Amazon Textract can be used in common business use cases.

Use case: Create smart search indexes

Your organization may have thousands, if not millions, of documents, records, and reports stored in archives. Finding what you need in such volume can be an arduous, manual process.

By using Amazon Textract, you can create a digital smart index of documents, searchable by key terms or key-value pairs. You can augment the data in the index by using natural language processing, an ML technique that can help you find insights and relationships in text. For example, automatically identify the language of the text, extract key phrases, places, people, brands, or events, and understand positive or negative sentiment to make the search index even more relevant.

Sample business applications:



In finance, banks or insurers can create smart search indexes to quickly find a specific customer name, or locate a record with specific parameters (such as all loans with x% rate).



In healthcare, parties managing a clinical trial can use Amazon Textract to gather the text from documents, use Amazon Comprehend Medical to understand that text, then build a searchable database—which can be used to identify patients that fit trial criteria.



In government or the public sector, institutions can create indexes to easily find reports and research on a specific public policy.

CUSTOMER TESTIMONY:



“Roche’s NAVIFY decision support portfolio provides solutions that accelerate research and enable personalized healthcare. With petabytes of medical PDF documents being generated in hospital systems every day, we needed a document extraction service to handle documents that have no standardized format. Amazon Textract provides the functionality to help us extract text from medical documents, so that we can then apply Natural Language Processing (NLP) to build a comprehensive, longitudinal view of patients, and enable both decision support and population analytics.”

Ram Balasubramanian,
Sr. Director of Software Engineering,
Roche Diagnostics Information Solutions

Use case: Build automated document workflows

Processing paperwork en masse can take extraordinary amounts of time and effort, with discrete pieces of information needing to go to various systems or teams (for example, specific data on a mortgage application needs to go to an HOA team or credit reporting system).

Amazon Textract provides you with the ability to automatically process forms and extract information from business documents, decreasing the level of human intervention necessary. When your business requires it, Amazon Textract offers the ability to combine the efficiency of artificial intelligence with human judgment, so you can make complex decisions with both speed and confidence.

Sample business applications:



Across industries, Amazon Textract can be used to automate common processes.

For example, financial institutions and lenders can automate loan applications by using the information contained in documents to initiate all of the necessary background and credit checks to approve the loan—so that customers can get instant results of their application rather than having to wait several days for manual review and validation.

Likewise, insurers can rapidly process claims and forms; and government agencies can automate applications such as driver's licenses and/or identification cards.

CUSTOMER TESTIMONY:



"At Cox Automotive, we are looking to transform the way the world buys, sells, owns and uses cars. To further modernize our automotive solutions, we will be leveraging Amazon Textract to accelerate how quickly cars can be transacted. With Amazon Textract, we can automatically capture and validate data from documents and forms, such as loan applications or vehicle titles, so decisions can be made more quickly. This will reduce customer effort and further streamline the process for everyone involved from the manufacturer to the buyer."

Bryan Landerman,
Chief Technology Officer,
Cox Automotive

Use case: Analyze trends in data over time

Today, businesses get smarter and more efficient by understanding patterns and trends across their company over time. A lot of that valuable data is locked in physical documents containing forms or tables. Traditional OCR methods can only pull out text, making it hard to understand the structure and logic of the data.

Amazon Textract automatically identifies form labels and values and extracts information from tables without compromising the structure, so you can understand patterns over time across millions of documents.

Sample business applications:



- Forecasting and running projections
- Evaluating sentiment over time
- Understanding how customer needs change

CUSTOMER TESTIMONY:

CHANGE HEALTHCARE

Change Healthcare is a leading independent healthcare technology company that provides data and analytics-driven solutions to improve clinical, financial, and patient engagement outcomes in the U.S. healthcare system.

“At Change Healthcare, we believe that we can make healthcare affordable and accessible to all by improving the timeliness and quality of financial and administrative decisions. This can be achieved by the power of ML technology to understand more from our data. But unlocking the potential of this information can often be difficult as it’s siloed in tables and forms that traditional optical character recognition (OCR) hasn’t been able to analyze. Amazon Textract further advances document understanding with the ability to retrieve structured data in addition to text, and now with the service becoming HIPAA compliant, we’ll be able to liberate the information from millions of documents and create even more value for patients, payers, and providers.”

Nick Giannasi, EVP and Chief AI Officer
Change Healthcare

Use case: Robotic process automation

Amazon Textract is a key component in business process automation workflows, and integrates into the emerging field of Robotic Process Automation (RPA).

Previous RPA solutions were akin to identifying objects in screenshots and taking pre-programmed automated steps with the identified objects. With machine learning now integrated, such as the photometric and geometric capabilities of Amazon Textract, the way information is captured has changed. Character recognition is smarter, and solutions like Amazon Textract can turn complex forms and tables into actionable, searchable data.

This means business processes can be automated and made more agile in a virtualized environment.

Sample business applications:



Sample business applications include culling data from documents, then using the insights in business or operational systems. Also, businesses can use Amazon Textract to enhance security in customer authentication processes or fraud detection.

REAL BUSINESS APPLICATIONS: UIPATH AND BLUE PRISM

"Amazon Textract will further differentiate UiPath's robotic process automation platform by enhancing UiPath's document understanding capabilities, enabling our customers to unlock critical business data from documents, transform that data into actionable business insights, and deliver those insights into line-of-business and operational systems."

Param Kahlon, Chief Product Officer
UiPath

"Blue Prism's connected-RPA can automate and perform mission-critical processes, allowing customers the freedom to focus on more creative, meaningful work. By using Amazon Textract, we've given our digital workforce another powerful tool for automation. Amazon Textract accurately analyzes data from various document types using machine learning, which enhances the digital transformation journey for our customers. Using additional AWS AI services like Amazon Comprehend and Amazon Rekognition, we can tackle challenges from added secure customer authentication processes to fraud detection capabilities. The intelligence and flexibility of Amazon Textract's form data extraction can elevate OCR to new levels in industries like financial services, retail, manufacturing and transportation to name a few."

Dave Moss, CTO and Co-Founder
Blue Prism



How to get started

Amazon Textract can help you extract text, automate workflows, and maintain compliance—eliminating manual effort and lowering processing costs. Once that text is captured, you can then take action on it within your business applications and add value to your business.

Learn more about Amazon Textract >>

Work with the AWS Partner Network (APN)

APN Partners have proven expertise in delivering document management services that leverage AWS machine learning and AI, including Amazon Textract. Partner solutions include OCR in Robotic Process Automation, system integration, platforms, and consultation.



© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.