

Amazon Textract

Easily extract text and data from virtually any document

Get started with Amazon Textract

Amazon Textract is a service that automatically extracts text and data from scanned documents. Amazon Textract goes beyond simple optical character recognition (OCR) to also identify the contents of fields in forms and information stored in tables.

Many companies today extract data from documents and forms through manual data entry that's slow and expensive or through simple optical character recognition (OCR) software that requires manual customization or configuration. Rules and workflows for each document and form often need to be hard-coded and updated with each change to the form or when dealing with multiple forms. If the form deviates from the rules, the output is often scrambled and unusable.

Amazon Textract overcomes these challenges by using machine learning to instantly "read" virtually any type of document to accurately extract text and data without the need for any manual effort or custom code. With Textract you can quickly automate document workflows, enabling you to process millions of document pages in hours. Once the information is captured, you can take action on it within your business applications to initiate next steps for a loan application or medical claims processing. Additionally, you can create smart search indexes, build automated approval workflows, and better maintain compliance with document archival rules by flagging data that may require redaction.