

Welcome to the OmniAI Document API!

How does it work?

To run document OCR with OmniAI, pass a **file** or **url** the **/extract** endpoint.

- Optionally pass in a desired JSON schema to extract structured output from that document. To run OCR without extracted response, simply omit the **schema** argument.
- This is an asynchronous api endpoint. The initial request will return a jobId.

Click Run 

Extract: **POST** /extract

Request Headers:

An API Key is required to access this endpoint.

Name	Value
x-api-key	your_api_key

Params

Either input File or URL is required but not more than one. See accepted file types below.

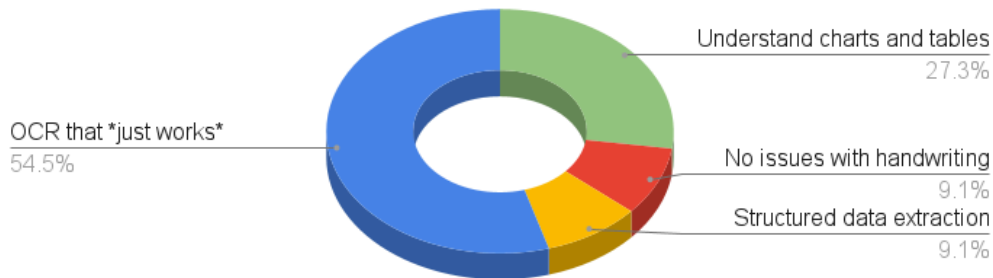
Name	Type	Description
url	string	URL of the document used for extraction
file	file	File used for extraction
schema	JSON (optional)	JSON schema
maintainFormat	boolean	Maintains format from the previous page. Defaults to false

```
curl -X POST https://api.getomni.ai/extract \  
  -H "x-api-key: my-api-key" \  
  -F "file=@path/to/your/file.pdf"
```

Turn any document into structured data

OCR is great! But the real magic comes from turning documents into beautiful structured JSON. Plug in a schema to the extract endpoint to return JSON output alongside your

Why do people love OmniAI?



Using a Schema

The **/extract** endpoint takes in an optional **schema** argument (example format below).

When passed in, the OCR result will be mapped to the specific JSON format you configured.

This is type safe. All responses will be validated against the input schema.

```
1 {
2   "type": "object",
3   "properties": {
4     "omniai_benefits": {
5       "type": "array",
6       "description": "Reasons to use OmniAI",
7       "items": {
8         "type": "object",
9         "properties": {
10          "benefit": {
11            "type": "string",
12            "description": "Section label from the chart"
13          },
14          "percentage": {
15            "type": "number",
16            "description": "Percentage on a scale of -1.00 to 1.00"
17          }
18        }
19      }
20    }
21  }
22 }
```

We can accept any of the following document types:

PDF's, Docx, PPT, and images are the most common document types, but we support just about everything. As long as we can convert the file to a visual format, it should work!

```
[
  "pdf", // Portable Document Format
  "doc", // Microsoft Word 97-2003
  "docx", // Microsoft Word 2007-2019
  "odt", // OpenDocument Text
  "ott", // OpenDocument Text Template
  "rtf", // Rich Text Format
  "txt", // Plain Text
  "html", // HTML Document
  "htm", // HTML Document (alternative extension)
  "xml", // XML Document
  "wps", // Microsoft Works Word Processor
  "wpd", // WordPerfect Document
  "xls", // Microsoft Excel 97-2003
  "xlsx", // Microsoft Excel 2007-2019
  "ods", // OpenDocument Spreadsheet
  "ots", // OpenDocument Spreadsheet Template
  "csv", // Comma-Separated Values
  "tsv", // Tab-Separated Values
  "ppt", // Microsoft PowerPoint 97-2003
  "pptx", // Microsoft PowerPoint 2007-2019
  "odp", // OpenDocument Presentation
  "otp", // OpenDocument Presentation Template
];
```

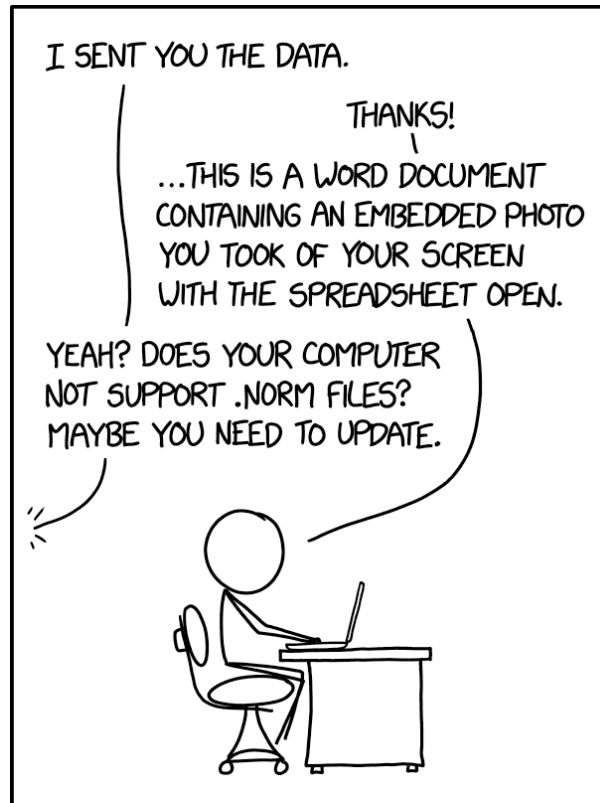
Flexible Pricing

Get started with our document extraction api. For more complex workflows, try out our platform for more complex workflows, or batch document processing.

API	Platform	Enterprise
\$500/month	\$2,000/month	Priced to fit your needs
Get access to:	Everything in API, plus:	Everything in Platform, plus:
Includes 10,000 pages per month	Includes 50,000 pages per month	Unlimited Documents
\$0.01 per page afterwards	Custom Document Pipelines	Connect to custom models
Structured data extraction	Batch requests	Fine Tuning
Up to 100 pages per document	Table automations	SSO / SAML
Up to 2 seats	Data Connectors	Export to Snowflake
	Export to Sheets	VPC Deployment
	No page limit on documents	Unlimited seats
	Up to 10 seats	

Why this problem?

The world seems determined to communicate via PDFs. We can't stop them, but we can make life a lot easier for everyone who has to work with them!



SINCE EVERYONE SENDS STUFF THIS
WAY ANYWAY, WE SHOULD JUST
FORMALIZE IT AS A STANDARD.