

PDFの文字を自動で読み取るしくみを学ぼう

「PDFに書かれた文章を、毎回手で書き写すのが大変…」そんな経験はありませんか？

今回は、Google Apps ScriptとGemini APIを使って、PDFの文字を自動で読み取り、
スプレッドシートに整理する方法を学びます。専門知識がなくても大丈夫。ゆっくり一緒に進めましょ
う。

手入力での作業



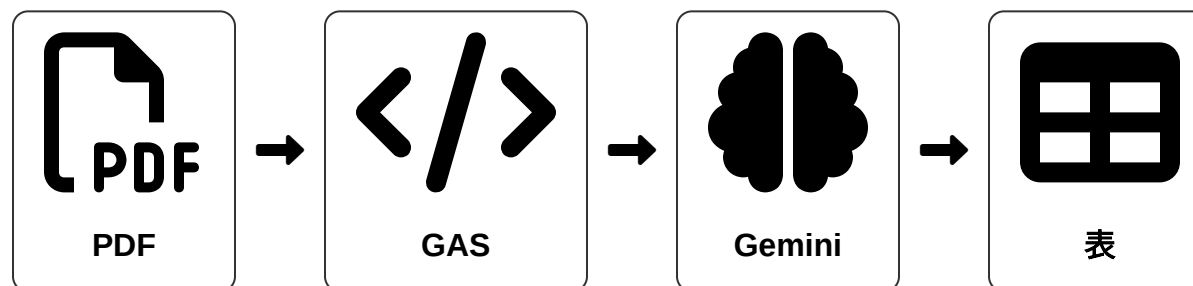
入力ミスが発生...

AIによる自動処理

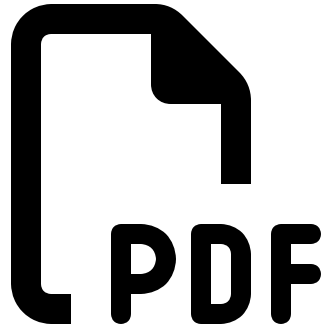
請求書データ			
請求番号	商品名	単価	税込
12345	商品A	2,500円	2,750円
12345	商品B	1,800円	1,980円
合計		4,300円	4,730円

正確に自動処理！

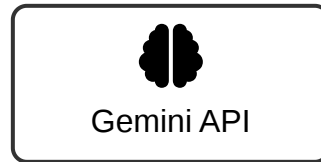
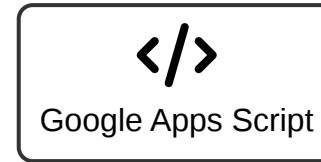
これから学ぶ内容



実例で学ぼう① どんなPDFでもAIで読み取れる？



今回は、領収書や請求書のようなPDFファイルを題材にします。どんな仕組みで文字を読み取るのか、まずは全体の流れをつかみましょう。



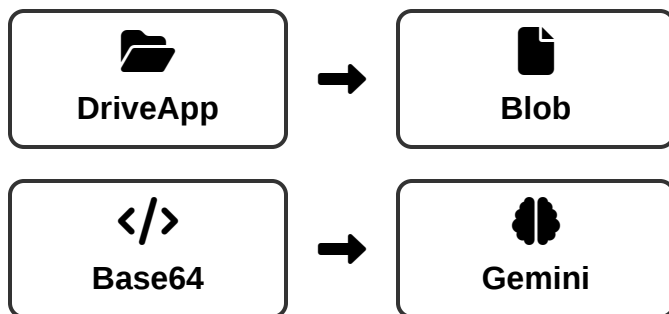
準備するもの

- | | | |
|---------------|-------------|---------------|
| ✓ Googleアカウント | ✓ PDFファイル | ✓ 演習用スプレッドシート |
| i 無料で利用できます | i 演習用サンプルあり | i テンプレート提供 |

実例で学ぼう② GASでPDFを扱うには？

Google Apps Scriptでは、PDFを一度テキスト形式に変換してGeminiに渡す方法が便利です。

コードのどの部分がどんな役割を果たすのか、図で見ながら理解しましょう。



```
// PDFファイルをDriveから取得
function getPdfFromDrive(fileId) {
  const file = DriveApp.getFileById(fileId);
  const blob = file.getBlob();
  return blob;
}

// BlobをBase64に変換
function convertToBase64(blob) {
  const base64Data = Utilities.base64Encode(
    blob.getBytes()
  );
  return base64Data;
}

// GeminiにPDFを送信して解析
function analyzeWithGemini(base64Data) {
  const apiKey = "YOUR_API_KEY";
```



コードはコピーすればそのまま使えます。プログラミング経験がなくても大丈夫！

実例で学ぼう③ Gemini APIで文字を抽出！

Gemini APIリクエスト

```
// PDFのBase64データをGemini APIに送信
const response = await gemini.generateContent({
  contents: [{
    parts: [{
      text: "PDFから情報を抽出してください。"
    }, {
      inlineData: {
        mimeType: "application/pdf"
```

Geminiからの応答

「請求書データを抽出しました：

- 会社名: 株式会社サンプル
- 請求金額: 55,000円
- 発行日: 2024年7月10日

抽出結果のスプレッドシート

田 請求書データ一覧

会社名	請求金額	発行日	支払期限
株式会社サンプル	55,000円	2024/7/10	2024/7/31
ABC商事	23,500円	2024/7/12	2024/7/25
山田電機	78,200円	2024/7/15	2024/8/10
テクノシステム	124,800円	2024/7/20	2024/8/20

どんな文書でも使える応用力



Gemini APIは請求書だけでなく、様々な形式の文書から情報を抽出できます。

手書き文字、複雑なレイアウト、表が含まれる文書でも高精度に読み取り可能です。

プロンプトを工夫することで、必要な情報だけを整理して抽出することができます。

実例で学ぼう④ スプレッドシートへの整形方法

📄 整形前のテキスト（ばらばら）

会社名：株式会社サンプル商事
請求書番号： INV-2025-0125
発行日： 2025年7月15日

商品名	数量	単価	金額
ノートパソコン	2	98,000円	196,000円
プリンター	1	45,000円	45,000円
モニター	3	25,000円	75,000円

小計： 316,000円
消費税： 31,600円
合計金額： 347,600円

📊 整形後のスプレッドシート

項目	内容
会社名	株式会社サンプル商事
請求書番号	INV-2025-0125
発行日	2025年7月15日

商品名	数量	単価	金額
ノートパソコン	2	98,000円	196,000円
プリンター	1	45,000円	45,000円
モニター	3	25,000円	75,000円

```
// スプレッドシートに整形して出力する関数 function
formatTextToSheet(extractedText) { const sheet =
SpreadsheetApp.getActiveSpreadsheet().getActiveSheet(); // 基本情報を抽出して表の上部に配置
const companyMatch = extractedText.match(/会社名: (.+)/);
const invoiceMatch = extractedText.match(/請求書番号: (.+)/);
const dateMatch = extractedText.match(/発行日: (.+)/); // 表形式にデータを整理
sheet.getRange("A1:B3").setValues([ ["項目", "内容"], ["会社名", companyMatch[1]], ["請求書番号", invoiceMatch[1]] ]); }
```

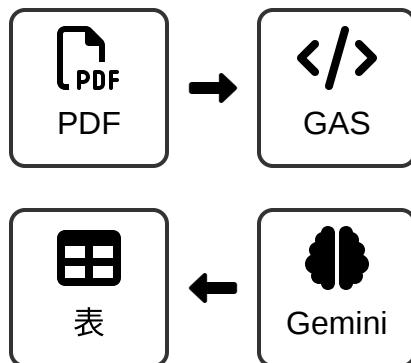


テキストから
構造化データへ変換

データの整理と
行列への自動配置

今日のふりかえりと活用アイデア

学習した内容



PDFから表形式データへの自動変換プロセス

活用アイデア

-  **請求書の整理**
毎月届く請求書PDFから金額や日付を自動抽出して管理表に反映
-  **アンケートの集計**
スキャンしたアンケート用紙からデータを自動で取り出して分析
-  **契約書の確認**
契約書PDFから重要項目を抽出して一覧管理
-  **議事録の整理**
PDFの議事録から決定事項やタスクだけを抽出して管理

今日できるようになったこと

- ✓ PDFからの文字抽出
- ✓ GASの基本的な使い方
- ✓ Gemini APIの活用
- ✓ スプレッドシートへの自動入力
- ✓ 表形式データの整形
- ✓ 業務自動化の基礎知識

やってみよう！自分のPDFを読み取ってみよう

あなたの職場で使えるとしたら？

演習手順

- 1 PDFファイルをアップロード**
演習用PDFをGoogle ドライブにアップロードしましょう
- 2 スクリプトをコピー&準備**
サンプルコードをGASエディタにコピーしてAPIキーを設定
- 3 スクリプトを実行**
実行ボタンを押して、PDFの文字読み取りを開始
- 4 結果を確認**
スプレッドシートに自動入力された結果を確認しましょう

あなたの業務で、どんな場面でPDFの自動読み取りが役立ちそうですか？

具体的な業務と、それによって解決できる課題を書き出してみましょう

※実際に試してみると、新しいアイデアも生まれるかもしれません

演習成功のポイント

- 💡 はじめは簡単なPDFから試してみましょう
- 📋 コードの意味を完全に理解する必要はありません
- 👥 困ったら周りの人と相談しながら進めましょう
- ★ 小さな成功体験が大きな自信につながります
- 🔄 一度動かせれば、応用はどんどん広がります
- ♥ 「できた！」という瞬間を大切にしましょう