أيوه، في حلول ممتازة على **Node.js** كمان تقدر تستخدمها لقراءة **النصوص من ملفات PDF** وأيضًا **عمل OCR على الصور داخل الملف**. خليني أشرح لك كيف تسوي نفس الشي هناك:

## **✅ الخطة:**

1. **نقرأ ملف PDF** باستخدام مكتبة مثل pdf-parse أو pdf-lib.
2. **نستخرج الصور من صفحات PDF**.
3. **نستخدم OCR** على الصور (باستخدام tesseract.js).

## **📦 المتطلبات:**

### **أولًا ثبّت البكجات:**

npm install pdf-parse tesseract.js fs path

## **✨ سكربت Node.js يقرأ النص + الصور:**

const fs = require("fs");  
const pdfParse = require("pdf-parse");  
const Tesseract = require("tesseract.js");  
  
const filePath = "file.pdf";  
  
fs.readFile(filePath, async (err, pdfBuffer) => {  
 if (err) throw err;  
  
 // 📝 قراءة النص العادي من PDF  
 const data = await pdfParse(pdfBuffer);  
 console.log("📝 النص المكتوب:\n", data.text);  
  
 // 📌 ملاحظة: pdf-parse ما يدعم استخراج الصور.  
 // تقدر تستخدم مكتبة مثل pdf-lib لاستخراج الصور من الصفحات  
 // وبعدين تمرر الصور لـ Tesseract  
  
 // مثال بسيط لاستخدام OCR على صورة:  
 Tesseract.recognize(  
 "image.png", // صورة تجريبية، لازم تكون مستخرجها من PDF  
 "ara+eng", // لو تبغى دعم عربي + إنجليزي  
 { logger: m => console.log(m) }  
 ).then(({ data: { text } }) => {  
 console.log("🖼️ النص داخل الصورة:\n", text);  
 });  
});

## **🚨 ملاحظة مهمة:**

* مكتبة pdf-parse سهلة لاستخراج النص، لكن ما تطلع الصور.
* لو تبغى تطلع الصور، استخدم مكتبة مثل:
  + [pdf-lib](https://pdf-lib.js.org/) (تقدر تقرأ الصور)
  + أو pdfjs-dist (نسخة JS من PDF.js تبع Firefox)

## **➕ لو تبي أجهز لك سكربت يقرأ النص + الصور من PDF بـ Node.js بشكل كامل، أقدر أضبطه لك مع pdf-lib و tesseract.js. تبيني أعمله؟**