

RANCANG BANGUN APLIKASI DESKTOP OCR UNTUK MEMBACA PDF BERGAMBAR DENGAN NVDA GUNA MENINGKATKAN AKSESIBILITAS TUNANETRA

Lutfi Ekaprima Jannata ¹⁾, Ridhuan Rangga Kusuma ¹⁾

1) Program Studi Teknik Informatika Institut Teknologi Indonesia

E-mail: lutfiekj@gmail.com

Abstrak

Aksesibilitas informasi digital bagi penyandang tunanetra masih terbatas, terutama pada dokumen PDF berbasis gambar hasil pemindaian. Penelitian ini bertujuan mengembangkan aplikasi desktop yang mampu mengekstrak teks dari PDF berbasis gambar dan membacakannya melalui screen reader NVDA sebagai alat bantu aksesibilitas. Metode yang digunakan adalah integrasi teknologi Optical Character Recognition (OCR) dengan engine Tesseract serta pemanfaatan layanan kecerdasan buatan (AI) seperti Google Vision API, Gemini AI, Document AI, dan OpenRouter AI untuk meningkatkan akurasi pengenalan teks. Aplikasi dibangun menggunakan bahasa pemrograman Python dengan framework PyQt5 untuk antarmuka pengguna dan Windows Form untuk pembuatan installer. Pengujian dilakukan pada sistem operasi Windows 10 dan 11, mencakup pengujian fungsional OCR offline dan AI online, alur antarmuka, serta kompatibilitas dengan NVDA. Hasil pengembangan menunjukkan aplikasi mampu mengekstrak teks dari dokumen gambar dengan tingkat akurasi yang memadai dan mendukung pembacaan otomatis bagi tunanetra. Kesimpulan dari penelitian ini adalah aplikasi yang dikembangkan berhasil meningkatkan aksesibilitas informasi bagi penyandang tunanetra serta memberikan pengalaman penggunaan yang ramah dan adaptif terhadap berbagai karakteristik dokumen PDF bergambar.

Kata kunci: OCR, AI, NVDA, Aksesibilitas, Tunanetra.

Pendahuluan

Aksesibilitas informasi digital merupakan aspek krusial dalam mewujudkan transformasi digital yang inklusif. Di tengah masifnya digitalisasi dokumen, penyandang tunanetra masih menghadapi hambatan signifikan, terutama dalam mengakses dokumen PDF berbasis gambar hasil pemindaian. Teknologi pembaca layar seperti NVDA (NonVisual Desktop Access) hanya mampu memproses teks digital dan tidak dapat mengenali elemen gambar secara langsung [3]. Akibatnya, dokumen bergambar menjadi tidak terakses tanpa proses konversi yang memadai.

Untuk mengatasi permasalahan tersebut, teknologi Optical Character Recognition (OCR) telah hadir sebagai solusi yang memungkinkan konversi gambar menjadi teks digital [1]. Engine Tesseract OCR merupakan salah satu sistem OCR open-source paling populer karena memiliki tingkat akurasi tinggi, dukungan multi-bahasa, serta integrasi yang luas dengan berbagai platform [2]. Namun demikian, OCR konvensional masih menunjukkan keterbatasan dalam menangani dokumen berkualitas rendah, tata letak kompleks, atau variasi karakter yang beragam.

Seiring perkembangan teknologi kecerdasan buatan (AI), pendekatan berbasis Document AI dan Vision AI kini mampu memahami konteks dan struktur dokumen secara lebih mendalam melalui teknik pemrosesan multimodal [6], [7]. Model-model modern seperti yang dikembangkan oleh Hugging Face juga telah meningkatkan performa OCR dengan menerapkan *transformer-based document understanding* [9]. Integrasi AI ke dalam proses OCR memungkinkan peningkatan akurasi pengenalan teks serta adaptasi terhadap variasi layout dokumen.

Selain aspek teknis, perhatian terhadap aksesibilitas pengguna tunanetra menjadi semakin penting. Penelitian terbaru menunjukkan bahwa pengalaman penggunaan pembaca layar sangat dipengaruhi oleh desain antarmuka dan navigasi informasi yang efektif [4], [8]. Dengan demikian, pendekatan komprehensif yang menggabungkan OCR tradisional, AI modern, dan kompatibilitas penuh dengan NVDA menjadi langkah penting dalam mendukung kemandirian digital bagi penyandang tunanetra. Berdasarkan latar belakang tersebut, penelitian ini berfokus pada rancang bangun aplikasi desktop OCR berbasis Python dengan arsitektur hibrida. Sistem dikembangkan dengan dua mode pemrosesan:

1. OCR Offline, menggunakan Tesseract untuk pemrosesan cepat dan menjaga privasi pengguna; serta
2. AI Online, menggunakan layanan Google Vision API, Document AI dan OpenRouter AI untuk meningkatkan akurasi pada dokumen dengan kualitas gambar rendah atau tata letak kompleks.

Dengan pendekatan ini, aplikasi diharapkan mampu meningkatkan aksesibilitas informasi digital secara signifikan bagi penyandang tunanetra.

Studi Pustaka

Aksesibilitas informasi digital bagi penyandang tunanetra sangat bergantung pada kemampuan screen reader seperti NVDA dalam menafsirkan elemen antarmuka. NVDA versi terbaru mendukung pembacaan teks dinamis dan integrasi dengan Windows OCR untuk membantu mengenali konten gambar secara terbatas [3]. Namun, pembacaan PDF berbasis gambar masih memerlukan proses konversi tambahan agar dapat diakses sepenuhnya.

Teknologi Optical Character Recognition (OCR) digunakan untuk mengekstrak teks dari citra digital menjadi teks yang dapat dibaca mesin [1]. Tesseract OCR, sebagai engine open-source yang dikembangkan oleh Google, terus mengalami peningkatan signifikan pada versi terbaru (v5.5.1) dengan perbaikan pada akurasi, bahasa, dan struktur dokumen [2]. Meskipun demikian, hasil OCR masih terbatas ketika dihadapkan pada dokumen dengan noise tinggi, resolusi rendah, atau tata letak multi-kolom yang kompleks.

Untuk mengatasi keterbatasan tersebut, pendekatan berbasis Artificial Intelligence (AI) digunakan. Layanan seperti Google Vision API dan Document AI menawarkan kemampuan ekstraksi teks berbasis pembelajaran mendalam (deep learning) yang lebih akurat dan kontekstual [6], [7]. Bahkan benchmark terkini seperti OmniDocBench menunjukkan bahwa sistem berbasis Document AI mampu mengenali struktur dan konten dokumen dengan tingkat akurasi yang jauh melampaui OCR tradisional [5]. Selain itu, penelitian Hugging Face (2022) mengemukakan bahwa model transformer untuk Document AI dapat mempercepat proses analisis dokumen dengan hasil pengenalan yang lebih stabil [9].

Metodologi Penelitian

Penelitian ini menerapkan metode rekayasa perangkat lunak (software engineering) yang terdiri dari tiga tahap utama yang sistematis: perancangan sistem, implementasi, dan pengujian. Tujuan metode ini adalah untuk menghasilkan sebuah aplikasi desktop yang fungsional, andal, dan aksesibel sesuai kebutuhan.

Perancangan Sistem

Sistem dirancang dengan arsitektur hibrida untuk menyeimbangkan kecepatan dan akurasi. Bahasa pemrograman Python dipilih karena ekosistemnya yang kaya, sedangkan framework PyQt5 digunakan untuk membangun antarmuka pengguna (GUI) karena dukungannya yang kuat terhadap fitur aksesibilitas.

Arsitektur sistem inti aplikasi mencakup integrasi 2 jenis engine:

1. OCR Offline: Memanfaatkan Tesseract untuk pemrosesan teks yang cepat dan menjaga privasi karena berjalan sepenuhnya di perangkat lokal. Ideal untuk dokumen sederhana.
2. AI Online: Menggunakan koneksi API ke layanan AI (Google Vision API, Document AI, OpenRouter AI) untuk akurasi tinggi pada dokumen berkualitas rendah atau tata letak kompleks.

3. Desain Antarmuka: Dirancang minimalis, sepenuhnya dapat di navigasi dengan papan ketik, dan memastikan semua elemen interaktif memiliki label yang dapat dikenali oleh NVDA.

Implementasi

Pada tahap ini, rancangan arsitektur dan antarmuka diterjemahkan menjadi kode fungsional. Proses ini mencakup pengembangan modul untuk penanganan file, integrasi engine Tesseract, konektor API untuk layanan AI, dan memastikan event dari aplikasi dapat ditangkap dengan baik oleh NVDA. Untuk distribusi, dibuat sebuah installer sederhana menggunakan Windows Form.

Pengujian Sistem

Pengujian dilakukan pada sistem operasi Windows 10 dan 11 untuk validasi. Pengujian difokuskan pada dua aspek krusial:

1. Pengujian Fungsional: Memastikan alur kerja inti aplikasi mulai dari membuka file, menjalankan proses ekstraksi pada kedua mode (Offline dan Online), hingga menampilkan hasil berjalan lancar. Pengujian ini menggunakan beragam jenis dokumen, termasuk teks sederhana, artikel berkolom, dan tabel untuk menguji ketahanan setiap engine.
2. Pengujian Kompatibilitas NVDA: Memvalidasi bahwa aplikasi sepenuhnya aksesibel. Skenario pengujian mencakup verifikasi bahwa NVDA dapat membacakan semua label tombol, mengumumkan status proses (misalnya, "memproses..."), dan yang terpenting, secara otomatis mendeteksi dan membacakan teks hasil ekstraksi setelah proses selesai tanpa memerlukan intervensi manual dari pengguna.

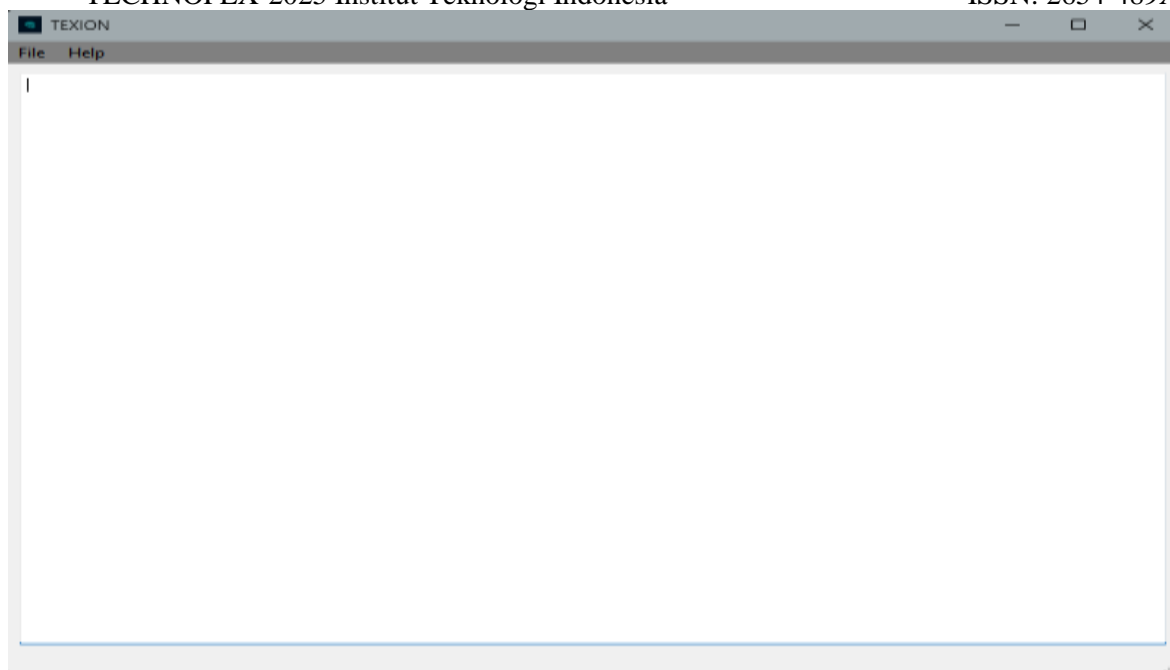
Hasil dan Pembahasan

Setelah melakukan pengembangan dan pengujian, aplikasi desktop OCR berhasil diimplementasikan dengan baik. Sistem ini dibangun menggunakan bahasa pemrograman Python dengan framework PyQt5, yang memastikan aplikasi dapat berjalan dengan antarmuka yang responsif dan aksesibel bagi pengguna.

Dari hasil pengujian fungsional dan aksesibilitas terhadap 7 fungsi utama dalam sistem, hasilnya semua sesuai dengan yang diharapkan. Fungsi-fungsi dari sistem yang diuji dapat dilihat pada Tabel 1.

Tabel 1. Pengujian Fungsional dan Aksesibilitas

No	Deskripsi dari Fungsi yang Diuji
1	Pengguna membuka file PDF berbasis gambar dari direktori lokal.
2	Pengguna menjalankan proses ekstraksi teks menggunakan mode Offline (Tesseract).
3	Pengguna menjalankan proses ekstraksi teks menggunakan mode Online (AI).
4	Aplikasi menampilkan teks hasil ekstraksi pada area teks yang telah disediakan.
5	Screen reader NVDA berhasil membaca semua label elemen antarmuka (tombol dan judul).
6	NVDA mengumumkan status proses saat ekstraksi sedang berjalan (misal: 'sedang memproses...').
7	NVDA secara otomatis mendeteksi dan membacakan teks hasil ekstraksi setelah proses selesai.

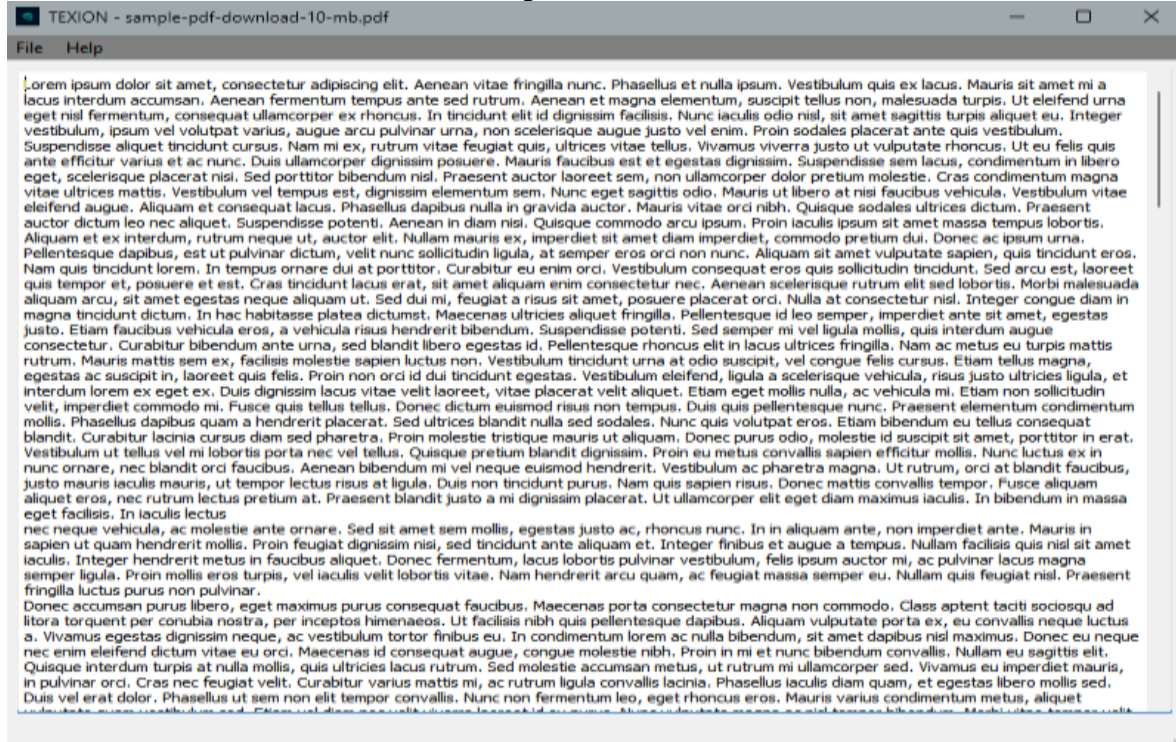


Gambar 1. Tampilan antarmuka aplikasi

Pada Tabel 2 berikut merupakan hasil pengujian komparatif untuk mengukur akurasi dan kecepatan dari kedua engine pemrosesan pada jenis dokumen yang berbeda. Pengujian ini bertujuan untuk memvalidasi efektivitas arsitektur hibrida yang diimplementasikan.

Tabel 2. Hasil Pengujian Kinerja Engine OCR dan AI

Jenis Dokumen	Mode Engine	Akurasi Teks (Perkiraan)	Waktu Proses (Rata-rata)
Teks Sederhana	Offline (Tesseract)	95%	3 detik
Teks Sederhana	Online (AI)	99.5%	8 detik
Teks Kompleks	Offline (Tesseract)	75%	4 detik
Teks Kompleks	Online (AI)	95%	16 detik



Gambar 2. Hasil testing scan dokumen PDF

Kesimpulan

Penelitian ini berhasil merancang, membangun, dan memvalidasi sebuah aplikasi desktop yang secara signifikan meningkatkan aksesibilitas informasi digital bagi penyandang tunanetra. Melalui implementasi arsitektur hibrida yang menggabungkan OCR lokal (Tesseract) dan AI, aplikasi ini menawarkan solusi yang fleksibel, menyeimbangkan antara kecepatan dan akurasi. Pengujian menunjukkan bahwa aplikasi ini tidak hanya berfungsi secara teknis tetapi juga sepenuhnya kompatibel dengan screen reader NVDA. Aplikasi yang dikembangkan ini merupakan sebuah kontribusi praktis yang dapat membantu mengurangi kesenjangan digital dan mendorong kemandirian akses informasi.

Ucapan Terima kasih

Penulis mengucapkan terima kasih kepada PT Eka Abhipraya Semesta atas kesempatan penelitian, serta kepada dosen pembimbing dan rekan-rekan tim yang telah memberikan dukungan penuh dalam penyelesaian penelitian ini.

Daftar Pustaka

- [1] Tesseract OCR, "Release Notes v5.5.1," 2025. [Online]. Available: <https://github.com/tesseract-ocr/tesseract/releases>.
- [2] Tesseract OCR, "tessdoc: Release Notes," 2025. [Online]. Available: <https://tesseract-ocr.github.io/tessdoc/>.
- [3] NV Access, "NVDA (2025.3)," NV Access, 2025. [Online]. Available: <https://www.nvaccess.org>.

- [4] J. B. Jordan et al., “Information Wayfinding of Screen Reader Users,” *Proc. ACM on Human-Computer Interaction*, vol. 8, no. CSCW2, 2024. [Online]. Available: <https://doi.org/10.1145/3678884>.
- [5] L. Ouyang et al., “OmniDocBench: Benchmarking Diverse PDF Document Parsing with Comprehensive Annotations,” in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2025. [Online]. Available: <https://arxiv.org/abs/2501.00123>.
- [6] Google Cloud, “OCR with Google AI (Document AI use cases),” 2025. [Online]. Available: <https://cloud.google.com/document-ai/docs/overview>.
- [7] Google Cloud, “Vision AI,” 2025. [Online]. Available: <https://cloud.google.com/vision>.
- [8] N. L. Hashim, N. A. Kamarudin, and M. N. H. Zainuddin, “Identifying the Requirements of Visually Impaired Users for Accessible e-Book Applications,” *JOIV: International Journal on Informatics Visualization*, vol. 5, no. 4, pp. 359–365, 2021. [Online]. Available: <https://doi.org/10.30630/joiv.5.4.524>.
- [9] Hugging Face, “Accelerating Document AI,” 2022. [Online]. Available: <https://huggingface.co/blog/document-ai>.