



Tungsten

OmniPage™ OCR:

Accuracy, Performance & AI

A Competitive Comparison



WHITE PAPER

TUNGSTEN
AUTOMATION

OmniPage OCR: An end-to-end OCR extraction offering

This white paper provides an in-depth comparison of **Tungsten Automation OmniPage™**, the industry-leading optical character recognition (OCR) solution against the leading Open-Source OCR offering. It highlights OmniPage's superior accuracy, performance (throughput), and pre- and post-processing functionality to provide a complete end-to-end OCR solution.

Key Topics Covered:

- OmniPage OCR accuracy compared to Open-Source OCR.
 - OmniPage's superior speed and performance.
 - Advanced preprocessing and document output options
-



Testing Overview

OmniPage offers three operational modes, enabling partners to tailor solutions to meet specific business requirements. For this analysis, testing was conducted using two of these modes:

- **Fast Mode:** Optimized for speed, with a slight trade-off in accuracy.
- **Accurate Mode:** Prioritizes precision for the most accurate results.
- **Balanced Mode:** Combines speed and accuracy for optimal performance
 - **Note:** Balanced Mode data not included but available on request

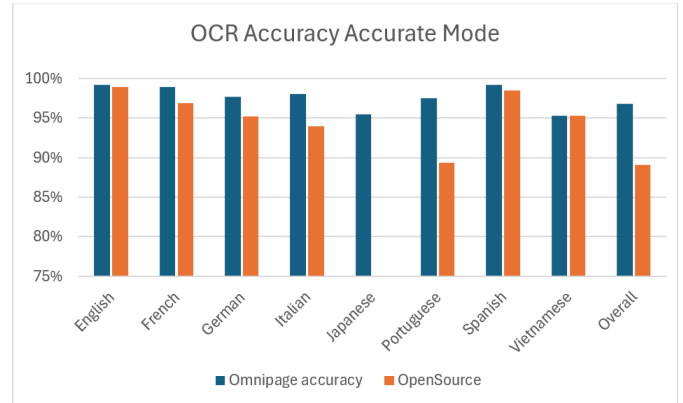
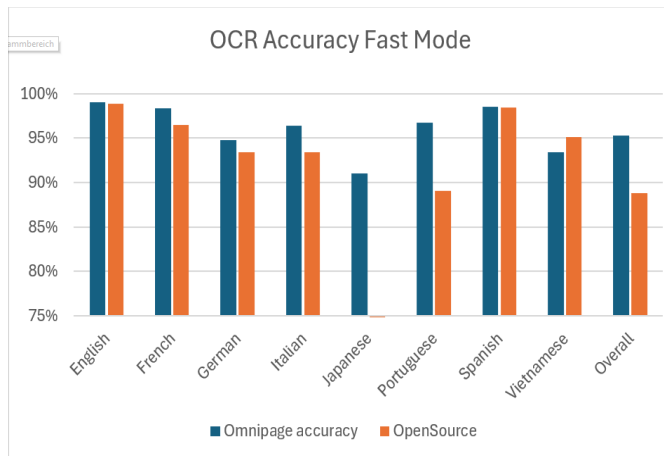
OmniPage regularly benchmarks our technology against competitors. The results consistently demonstrate OmniPage's superior OCR accuracy and performance, reaffirming its position as a market leader.

In addition to performance, we will review the OmniPage Image Processing algorithms and document output formats that provide developers a complete "ingest to output" toolkit offering.

OCR Accuracy

When comparing the **OmniPage Capture Software Development Kit (CSDK)** to the leading Open-Source OCR solutions, accuracy emerges as the most critical feature for Product Owners and Developers to decide on which technology to utilize. OmniPage consistently outperforms Open-Source offerings. In this white paper, we will compare OmniPage versus a leading competitor in two modes, **Fast Mode** and **Accurate Mode**:

Our testing was done in 28 languages (OmniPage provides recognition for over 130 languages). Our charts list the top eight demanded languages. Additional language comparisons are available on request.



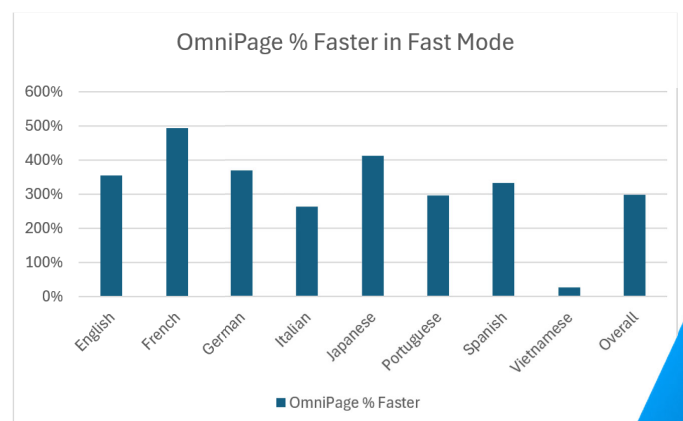
- **Fast Mode:** OmniPage is more than **97% accurate**. **Open-Source software produces 1.8x as many errors** as OmniPage (on average) when processing Western documents in Fast Mode.
- **Accurate Mode:** OmniPage is **98% accurate**. **Open-Source software produces 2.7x as many errors** as OmniPage (on average) when processing Western documents in Accurate Mode.

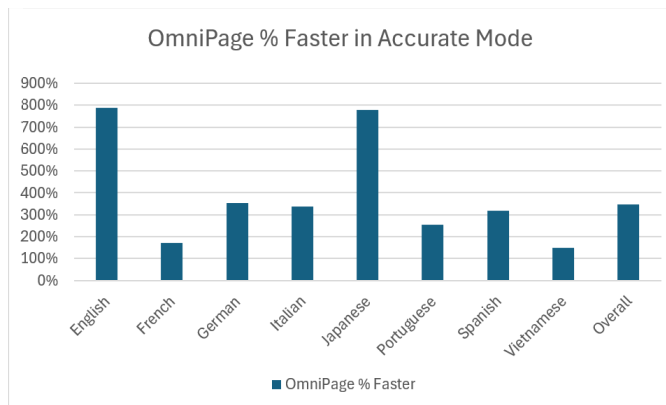
While these percentages may seem small, even minor improvements in accuracy significantly reduce errors that lead to exception handling, saving time and resources plus increases the ability to find (via search) information on the documents.

Performance (Speed Improvements)

Performance is developers' second most requested feature, particularly for desktop applications where users expect instantaneous results. Open-Source offerings are not optimized for performance, and our testing has found a very large gap between the time it takes OmniPage to process a document versus the leading Open-Source offering.

OmniPage excels in reducing processing time, delivering faster results in both Fast and Accurate modes:





- **Fast Mode:** OmniPage is **3 times faster** and up to 5x in some languages.
- **Accurate Mode:** OmniPage is **over 3 times faster** and up to 8x in some languages.

Faster processing equates to less computing power (CPUs) required for processing, reducing the cost of processing.

The combination of speed and accuracy makes OmniPage the ideal OCR solution for businesses seeking efficiency and reliability.



Image Preprocessing and Output Options

Similar to a high-end Swiss Army knife, having everything in one solution provides:

- Seamless, interconnected integration
- One code set to manage
- One company to contact for your development and support
- One set of libraries to upgrade when enhancements are available.

Two other valuable benefits that OmniPage offers over the leading Open-Source OCR are:

Enhanced Image Preprocessing

OmniPage's preprocessing algorithms, refined over 20 years, ensure high-definition images for optimal OCR results.

Key capabilities include:

- **Resolution Enhancement:** Adjusts images to the ideal 200-300 DPI range.
- **Binarization:** Converts color or gray-scale images to black and white for OCR processing.

- **Auto-Rotation and Deskewing:** Automatically corrects image orientation.
- **Adaptive Noise Removal:** Eliminates specks and artifacts for cleaner images.

For example, the leading Open-Source technology can address documents with a 5-degree tolerance only for skewed scans, while OmniPage can handle 15-degree skew and can also straighten 3D skew (when an open book is captured with a digital camera).

There are Open-Source third-party options to address deskewing and the other image preprocessing features, but they require managing additional code sets that need to be learned, tweaked, and managed.

Having everything in one place simplifies the management of the complete set of tools required to manage the OCR process from start to finish.

Export options (Output formats)

The leading Open-Source offering provides five output options (txt, PDF, hocr, tsv and PDF with text layer only)

OmniPage offers **40 different output options**. The output formats requested the most from our 500+ customers are DocX, DocXEs-say Excel2000, Html50, PDF, PDFImageOn-Text, UFormattedTxt, UText, XlsX, and XML

OmniPage is designed to preserve the layout and formatting of the original document, including columns, tables, and graphics, in these output formats. This makes it particularly useful for creating editable and searchable documents while maintaining the original document's structure.

The most requested output type is PDF.

In this area, OmniPage delivers two additional features that every developer needs.

First, supercompression. With our advanced MRC compression, OmniPage can reduce any document to 10% of a standard PDF file, with some files compressed to 1% of a standard PDF file. With the move to Cloud solutions and the growing cost of uploading and downloading files, having a super compressed file reduces the cost of moving documents.

Second, we adhere to the latest PDF standards. The independent PDF organization continuously updates the PDF standard, and our PDF output complies with the latest PDF standards.



TAA Compliance

TAA compliance refers to the adherence to the Trade Agreements Act (TAA), a U.S. federal law which establishes guidelines for government procurement to ensure that products and services purchased by the U.S. government are sourced from either the United States or designated countries that have trade agreements with the U.S.

Benefits of TAA Compliance

- Access to Government Opportunities
- Risk Mitigation and Legal Protection
- Global Market Access via GATT and WTO
- Risk Prevention

Conclusion

Optical Character Recognition is a small piece of any solution that needs to extract data from an image, but it is an essential piece. Similar to tires on a car. Tires represent just a small piece of the overall investment, but the car can't drive without them.

In today's fast-paced, mobile, and remote work environments, businesses face increasing challenges in driving productivity and reducing costs. Paper documents often add complexity to workflows, making digital transformation essential.

OmniPage, a **TAA compliant OCR offering**, empowers organizations to streamline document processing with minimal programming effort, enabling teams to focus on higher-value tasks.

Whether your business processes a few documents or millions of pages daily, OmniPage provides the tools to enhance productivity and efficiency. Its powerful and versatile **Capture SDK** allows seamless integration of OCR and imaging capabilities into any platform.

Next Steps

Let us help your development team replace your existing OCR technology with OmniPage to achieve:

- Better OCR accuracy
- Faster performance
- A soup-to-nuts solution for all your OCR needs
- Pre- and post-sales support to reduce the time and effort to get your OCR project up and running and maximize your OCR results
- TAA Compliant offering



For more information, download the **Art and Science of OmniPage OCR** white paper [here](#) or contact us via email at jennifer.phillips@tungstenautomation.com.

