

# THÔNG TIN VỀ CÁC GIẢI PHÁP OCR NHẬN DẠNG KÝ TỰ TIẾNG VIỆT

*OCR = Optical Character Recognition = Nhận dạng ký tự quang học*

| TT | Sản phẩm - giải pháp | Tác giả (cá nhân / doanh nghiệp) + Website  | Mô tả các tính năng nhận dạng  | Khả năng tích hợp với hệ thống khác   | Khả năng hợp tác   |                                   |                      | Xếp hạng Việt Nam | Xếp hạng thế giới | Thông tin khác |
|----|----------------------|---|--|---|--|-----------------------------------|----------------------|-------------------|-------------------|----------------|
|    |                      |   |  |   | Bán license  | Bán license và giải pháp tích hợp | Bán toàn bộ mã nguồn |                   |                   |                |
| 1  | VnDOCR4              | Phòng nhận dạng & Công nghệ tri thức - Viện công nghệ thông tin.<br><a href="http://vndocr.com/home/">http://vndocr.com/home/</a> | <b>Môi trường:</b> PC với hệ điều hành Windows 2000, XP, NT, 7 (32Bit)...<br><b>Tiện ích:</b> Bộ gõ chữ Việt, bộ font ABC, VNI, Unicode,...<br><b>Thông tin đưa vào:</b> Quét trực tiếp các loại sách báo, văn bản qua máy quét (VnDOCR tương thích với tất cả các hãng Scanner HP, Kodak, Epson,...).<br>Đọc và xử lý hơn 30 dạng tệp tin ảnh phổ dụng nhất như PCX, BMP, TIF, GIF, JPG,...<br>Ảnh tài liệu quét -> VnDOCR nhận dạng chỉ yêu cầu: ảnh đen/trắng (B/W), độ phân giải 300 dpi.<br>Có thể nhận dạng trực tiếp tài liệu quét qua Scanner không cần lưu trữ dưới dạng tệp ảnh trung gian. Các trang tài liệu có thể được quét và lưu trữ dưới dạng tệp tin nhiều trang.<br><b>Kết quả:</b> Các tệp có định dạng: *.doc, *.xls, *.pdf,... | Hiện không có thông tin về việc cung cấp giải pháp tích hợp, cung cấp API cho các đối tượng doanh nghiệp phát triển ứng dụng.<br><br>Bán sản phẩm cài đặt có License cho các máy tính cá nhân hoặc tổ chức. | Hiện đang cung cấp phiên bản Client có license tính giá:<br>1: 2.200.000 VNĐ<br>2 – 5: 1.600.000 VNĐ<br>6 – 10: 1.400.000 VNĐ<br>11: 1.100.000 VNĐ | N/A                               |                      | 2                 | N/A               |                |

|   |   |  |   |   |   |  |  |   |     |   |
|---|---|--|---|---|---|--|--|---|-----|---|
| 2 | Adaptive Document Recognition Technology của hãng ABBYY | Sản phẩm của hãng ABBYY.<br><a href="http://www.abbyy.com/">http://www.abbyy.com/</a><br><b>ABBYY Partner tại Việt Nam (chính thức trên website của Abbyy):</b><br>Công ty Đầu tư và Phát triển Đông Kinh<br><a href="http://www.sohoatailieu.com.vn">http://www.sohoatailieu.com.vn</a> | Cung cấp 4 sản phẩm chính:<br>ABBYY FineReader 11.(Nhận dạng văn bản tiếng việt 99%).<br>ABBYY Recognition Server 3.0 (Giải pháp số hóa văn bản cho một doanh nghiệp).<br>ABBYY FlexiCapture 10 (Bóc tách văn bản tiếng việt)<br><b>ABBYY FineReader Engine 10.0</b><br>(Thư viện hỗ trợ việc xây dựng số hóa)  | Tích hợp với các ứng dụng thông qua bộ SDK của ABBYY.               | - Cung cấp dịch vụ trực tuyến<br>1Dcoin/1trang A4<br>(1Dcoin/1000VND) | - Hỗ trợ SDK<br>(Thông tin khoảng 20.000 \$)<br>- run time license |  | 1 | N/A | Giải pháp này có khả năng giữ được cả định dạng file.   |
| 3 | Phần mềm nhận dạng tiếng việt ABBYY                     | Sản phẩm của hãng ABBYY.<br><a href="http://www.abbyy.com/">http://www.abbyy.com/</a><br><b>Công ty phân phối:</b><br>Công ty CP Đầu Tư Thương Mại và Phát Triển Công Nghệ FSI<br><a href="http://sohoatailieu.com/">http://sohoatailieu.com/</a>  | <b>Đầu vào:</b> Hầu hết định dạng ảnh đầu vào như: PDF, TIFF, JPEG, GIF, PNG, BMP, PCX, DCX, DjVu...<br><b>Kết quả:</b> Nhận dạng lưu trữ dưới dạng MS Word, MS Excel, HTML, TXT, XML, PDF, PDF 2 lớp, trong đó định dạng PDF 2 lớp là một định dạng hoàn hảo cho việc lưu trữ và khai thác tài liệu. Với định dạng này, người đọc có thể đọc trung thực ảnh gốc nhờ lớp ảnh bên trên, các công cụ tìm kiếm có thể tìm kiếm toàn văn trên văn bản nhờ lớp text nhận dạng được bên dưới. | - Phân phối SDK của ABBYY --> Có khả năng tích hợp vào các ứng dụng | N/A   | Phân phối SDK:<br>8.000-12.000 \$                                  |  | 1 | N/A | -Đánh giá của FSI thì bộ SDK của ABBYY vẫn chưa đáp ứng được việc nhận dạng tiếng Việt.<br><b>FSI có giới thiệu một sản phẩm số hóa riêng của họ, tự đánh giá tốt hơn trong việc nhận dạng tiếng Việt.</b><br><b>Contact:</b> Mr. Thuận<br>0904.632.189 |

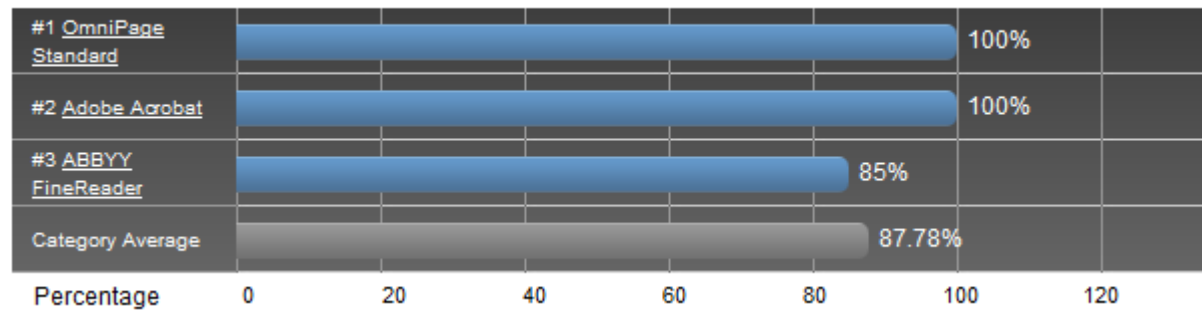
|   |                   |   |   |  |                      |                                     |  |     |     |   |
|---|-------------------|---|---|--|----------------------|-------------------------------------|--|-----|-----|---|
| 4 | VietOCR           | <a href="http://vietocr.sourceforge.net/">http://vietocr.sourceforge.net/</a> | <p><b>Nền tảng:</b> Dự án này được phát triển dựa trên nền tảng mã nguồn mở tesseract-ocr do Google tài trợ.</p> <p><b>Đầu vào:</b> Các file định dạng PDF, TIFF, JPEG, GIF, PNG, BMP</p> <p><b>Đầu ra:</b> Văn bản kiểu text. Độ chính xác khoảng 70-80%. Đối với các văn bản viết tay, độ chính xác không cao (đánh giá của cộng đồng mạng)</p> | - Một sản phẩm free sử dụng trên windows. Không có thông tin về API tích hợp.  | Sản phẩm free.       | N/A                                 |  | N/A | N/A |   |
| 5 | Laserfiche        | <a href="https://www.laserfiche.com">https://www.laserfiche.com</a>           | <p><b>Đầu vào:</b> PDF, TIFF, JPEG, GIF, PNG, BMP</p> <p><b>Kết quả:</b> MS Word, MS Excel, PPT.</p>  | <p>- Tích hợp thông qua SDK của laserfiche.</p> <p>- Đăng kí tài khoản và download SDK từ website của laserfiche</p> |                      | Cần có LSAP license để download SDK |  |     |     | <p>Năm 2009, TinhVan Group có kí hợp tác "Số Hóa dữ liệu" với Laserfiche thông qua Quadralink. Một số sản phẩm của Tinh Vân liên quan đến số hóa:</p> <ul style="list-style-type: none"> <li>- Phần mềm kiểm tra và chuyển đổi dữ liệu đa năng (iTransform)</li> <li>- Phần mềm sao lưu phục hồi dữ liệu TVSmartBackup</li> </ul> |
| 6 | OmniPage Standard | <a href="http://www.nuance.com/index.htm">http://www.nuance.com/index.htm</a> | <p><b>Môi trường:</b> Windows 32/64 bit</p> <p><b>Đầu vào:</b> images, pdf</p> <p><b>Đầu ra:</b> *.doc, *.xls, *.ppt, *.pdf, *.html</p>   | - Phần mềm đóng gói cho windows. Không có thông tin về tích hợp API.   | 150\$ cho 1 license. |                                     |  |     |     |   |

## SO SÁNH MỘT SỐ PHẦN MỀM OCR NỔI TIẾNG NHẤT

- OmniPage Standard
- Adobe Acrobat
- ABBYY FineReader

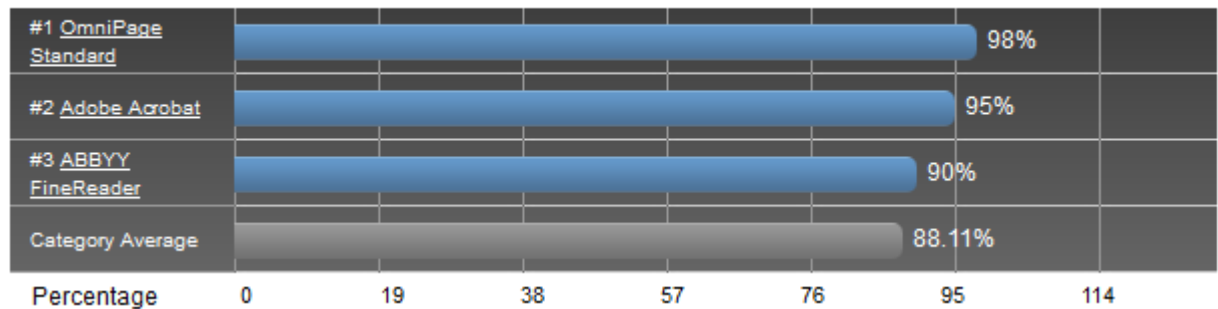
Tính dễ sử dụng:

Compare to the Top 3 OCR Software Review



Độ chính xác:

Compare to the Top 3 OCR Software Review



Ngôn ngữ hỗ trợ:

Compare to the Top 3 OCR Software Review

