# **Here are 5 common OCR (Optical Character Recognition) tools that are commonly used.**

1. **Tesseract**: Tesseract is an open-source OCR engine developed by Google.
   * Tesseract 4 adds a LSTM (Long Short-Term Memory) based OCR engine. But also supports the legacy OCR engines of tesseract 3.
   * Tesseract has Unicode (UTF-8) support.
   * Tesseract can recognize over 100 languages.
   * Tesseract supports various image formats including PNG, JPEG and TIFF.
   * Tesseract supports various output formats including plain text, hOCR (HTML), PDF, TSV etc.
   * Tesseract does not include a GUI application.
   * Tesseract can be trained to recognize other languages.
2. **ABBYY FineReader:** ABBYY FineReader is a commercial OCR software that can recognize and convert a wide range of languages and document formats.
   * Commonly used in business settings for document management and data extraction.
   * The program runs under Microsoft windows 7 or later and Apple macOS 10.12 Sierra or later.
   * Version 15 support text in 192 languages and has a built-in spell check for 48 of them.
   * The program allows users to compare documents, add annotations and comments and schedule batch processing.
   * FineReader recognizes new characters by training characters so that they are added to the recognition alphabet, selecting additional characters from a list and adding them to an alphabet of the selected language, and adding the domain specific vocabulary to the FineReader’s built-in lexicon.
3. **Amazon Textract**: Amazon Textract is a cloud-based OCR service provided by Amazon Web Services (AWS).
   * Users do not need to install or maintain the software locally and can access the service from anywhere with an internet connection.
   * It uses advanced machine learning models to accurately recognize and extract text from a variety of document types, including structured and unstructured documents. It can extract data from tables forms and handwriting.
   * It is highly scalable allowing users to process large volumes of data quickly and efficiently.
   * It can integrate with other Amazon web services to automate document processing workflows.
   * It provides easy simple and easy to use API for developers to integrate OCR functionality to their apps. It also includes pre-built connectors for popular business applications like SharePoint, making it easy to integrate it with existing workflows.
4. **OCRopus**: OCRopus is an open-source OCR system that includes machine learning algorithms for character recognition. It was developed by the German Research Center for Artificial Intelligence.
   * It is freely available and can be customized and extended by developers to meet their specific needs.
   * It supports multiple languages and character sets making it a versatile OCR solution.
   * It contains pretrained model on common languages and can be trained on new models and fonts.
   * In addition to the recognition scripts themselves, there are a number of scripts for ground truth editing and correction, measuring error rates, determining confusion matrices, etc.
   * The default parameters and settings of OCRopus assume 300dpi binary black-on-white images.
   * It is designed to integrate with other OCR and image processing tools, such as Tesseract OCR and ImageMagick.
5. **Readiris**: Readiris is a commercial OCR software that can recognize and convert scanned documents and PDF files into searchable and editable text.
   * It also includes document conversion features that allow users to convert documents into a variety of formats, including Word, Excel, and PDF/A.
   * It supports multiple languages and character sets, making it a versatile OCR solution for users worldwide. It includes pre-built recognition models for over 130 languages and can also be trained in new languages and fonts.
   * It includes advanced image processing features that can enhance the quality of scanned documents and images before OCR.
   * It can remove noise, skew, and other distortions, and can also improve the legibility of text and graphics.
   * It includes batch processing features that allow users to process multiple documents at once.
   * It can integrate with a variety of other tools and applications, including Microsoft Office, Adobe Acrobat, and cloud services such as Dropbox and Google Drive.