Challenges of OCR in Python  
  
With this vast number of applications, some drawbacks are associated with OCR that hampers its performance. When the input images are of degraded quality, it  
becomes challenging to perform Optical Character Recognition and extract text. Some common degradations are low lighting, less contrast between foreground and  
background, low-resolution images, broken or faded fonts, black noise due to repeated xerox, and so on. Multiple OCRs in the industry perform well, but out of all of  
them, Tesseract OCR is the most reliable. Tesseract is unique because it comprises various functionalities you can leverage to customize it for multiple tasks. So now,  
let's dive deep into the working of Tesseract OCR Engine, its features, and the various applications it can support.  
  
How does Tesseract OCR Python work?  
  
The tesseract library uses a defined set of techniques for Optical Character Recognition processing. First, the image is converted into binary then a connected  
component analysis takes place, which stores the outlines of the components. These outlines are then gathered as blobs. Next, blobs are structured into text lines which  
are further broken down into words based on the spacing of characters.  
  
Once the words are detected, the next phase is about recognition. Recognition is a two-step process. In the first step, each word is tried to be recognized, and  
satisfactory words are passed to an adaptive classifier. When everything is parsed once, a second pass is taken to recognize text that were not recognized in the first  
pass.  
  
Ultimately, the algorithm makes a final pass to resolve fuzzy spaces and locate small-cap text.  
  
The given flowchart describes all the steps taken by the Tesseract OCR mentioned above to provide the output result -  
==================================================  
Tesseract is one of the most widely used and open-source OCR tools developed by Hewlett-Packard (HP) Labs for the data science industry in the 1990s and later  
released as open-source in 2005. You can use it with an API to send requests and receive responses or use it directly. Tesseract is available in multiple languages; it  
supports more than 100 languages, including right-to-left languages. A significant advantage of Tesseract is that it is compatible with several programming languages  
through various wrappers. Along with the fundamental algorithms of text finding and line finding, the latest version of Tesseract also incorporates an Al-based approach  
(LSTM neural network) for better detecting and recognizing inputs with variable sizes.  
==================================================  
Filesystem Hierarchy Standard  
  
   
  
Article Talk  
  
   
  
From Wikipedia, the free encyclopedia  
  
The Filesystem Hierarchy Standard (FHS) is a reference describing the conventions used for the layout 0  
a UNIX system. It has been made popular by its use in Linux distributions, but it is used by other UNIX  
variants as well.'4! It is maintained by the Linux Foundation. The latest version is 3.0, released on 3 June  
2015.7)  
  
Directory structure [eit]  
  
Inthe FHS, all files and directories appear under the root directory / , even if they are stored on different  
physical or virtual devices. Some of these directories only exist on a particular system if certain subsystems  
such as the X Window System, are installed.  
  
   
  
Most of these directories exist in all Unix-like operating systems and are generally used in much the same  
way; however, the descriptions here are those used specifically for the FHS and are not considered  
authoritative for platforms other than Linux. -  
==================================================