# Tessreact-ocr 识别图片文字

## 环境准备

### 安装 leptonicha

1. 下载 wget <http://www.leptonica.org/source/leptonica-1.73.tar.gz>
2. 解压tar -xvf leptonica-1.73.tar.gz
3. cd leptonica-1.73
4. ./configure
5. make
6. make install

### 安装 ImageMagick

ImageMagick是什么？

ImageMagick是一个用于查看、编辑位图文件以及进行图像格式转换的开放源代码软件套装

我在这里之所以提到ImageMagick是因为某些图片格式需要用这个工具来转换。

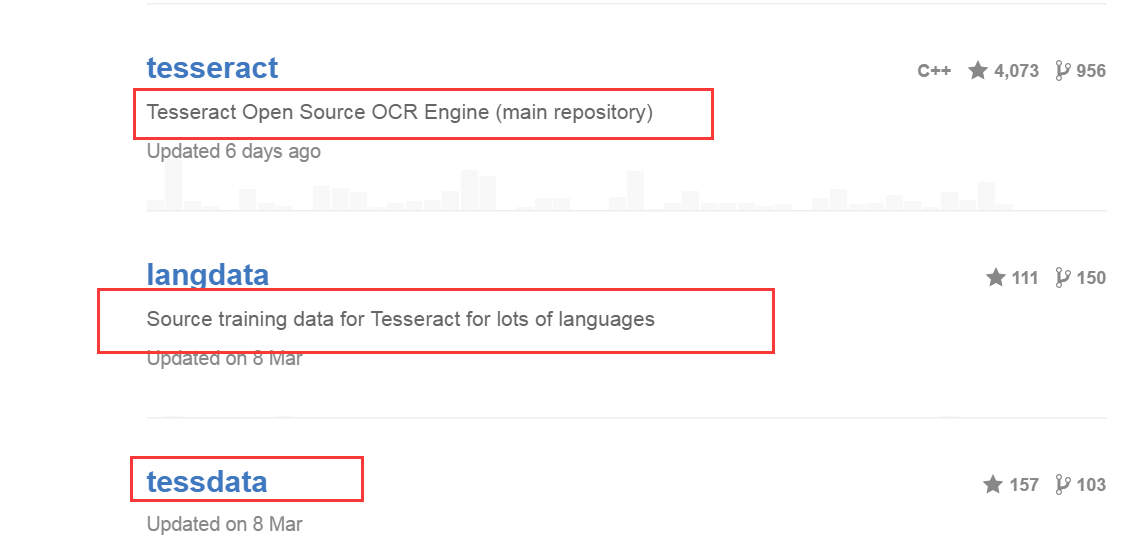
yum ImageMagick

### 安装 Tesseract

1. 下载 wget <https://github.com/tesseract-ocr/tesseract/archive/3.04.00.tar.gz>
2. tar -xvf 3.04.00.tar.gz
3. cd tesseract-3.04.00
4. ./autogen.sh
5. ./configure
6. make
7. make install
8. ldconfig

### 安装 Tessdata

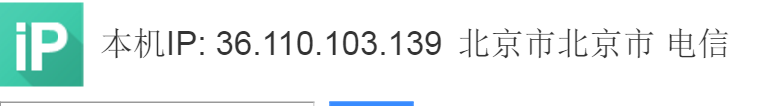
1. 下载训练库：wget <https://github.com/tesseract-ocr/tessdata/archive/master.zip>
2. unzip tessdata-master.zip
3. cp ./tessdata-master/\* /usr/local/share/tessdata



1. tesseract OCR引擎
2. langdata OCR 语言库
3. tessdata 直接存放在Linux操作系统的/usr/local/share/tessdata/

### 测试 Tesseract

测试图片：



命令：**tesseract tess4j.png tp1 -l eng+chi\_sim**

命令参数解释 图片名称 结果保存的文件名 -l表示语言lang eng表示英文 chi\_sim表示中文简体

多种语言使用“+” 连接

测试结果：

C:\Users\sere\AppData\Roaming\Tencent\Users\406214587\QQ\WinTemp\RichOle\W[YWH$@S($VC1T)[}V955YP.png

## Java版本实现参考：

<http://blog.csdn.net/lmj623565791/article/details/23960391>