

Tesseract Engine .net Wrapper – a quick overview

Author: Cong Nguyen

Release date: 2011/07/04

<http://code.google.com/p/tesseractdotnet/>

A. Engine

*This version has worked on the latest **tesseract-ocr engine (r590)**.*

*Integrated leptonica library into tesseract engine directly, so the tesseract assembly can load **without liblptxxx.dll**.*

a. Recognize

```
public:
    //
    // Recognize text from image file
    //
    String* Recognize(String* filePath);

    //
    // Recognize text from image
    //
    String* Recognize(System::Drawing::Image* image);

    //
    // Recognize text from binary image
    //
    String* RecognizeBinaryImage(unsigned char* binData, int width, int height);

    //
    // Recognize text from grey image (8-bits per pixel)
    //
    String* RecognizeGreyImage(unsigned char* greyData, int width, int height);

    //
    // Recognize text from grey image (depth: 8bpp, internal_format: 16 bits per pixel)
    //
    String* RecognizeGreyImage(unsigned short* greyData, int width, int height);
```

b. Analyze Layout

```

public:
    //
    // Initialize tesseract for page-layout analysis
    //
    bool InitForAnalysePage();

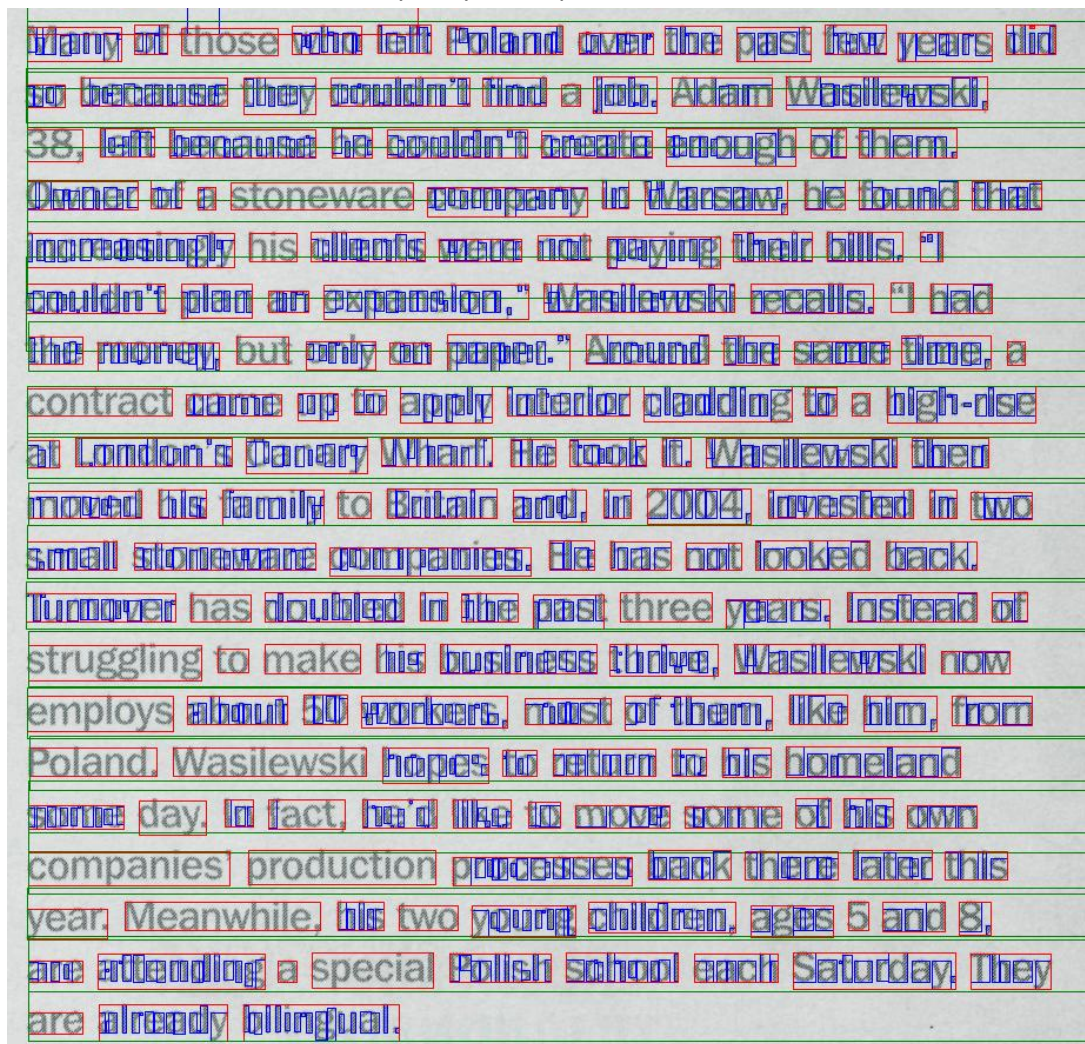
    //
    // Set page-layout segmentation mode
    //
    bool SetPageSegMode(ePageSegMode psm);

    //
    // AnalyseLayout interfaces
    //
    DocumentLayout* AnalyseLayout(System::Drawing::Image* image);
    DocumentLayout* AnalyseLayoutBinaryImage(unsigned char* binData, int width, int height);
    DocumentLayout* AnalyseLayoutGreyImage(unsigned char* greyData, int width, int height);
    DocumentLayout* AnalyseLayoutGreyImage(unsigned short* greyData, int width, int height);

```

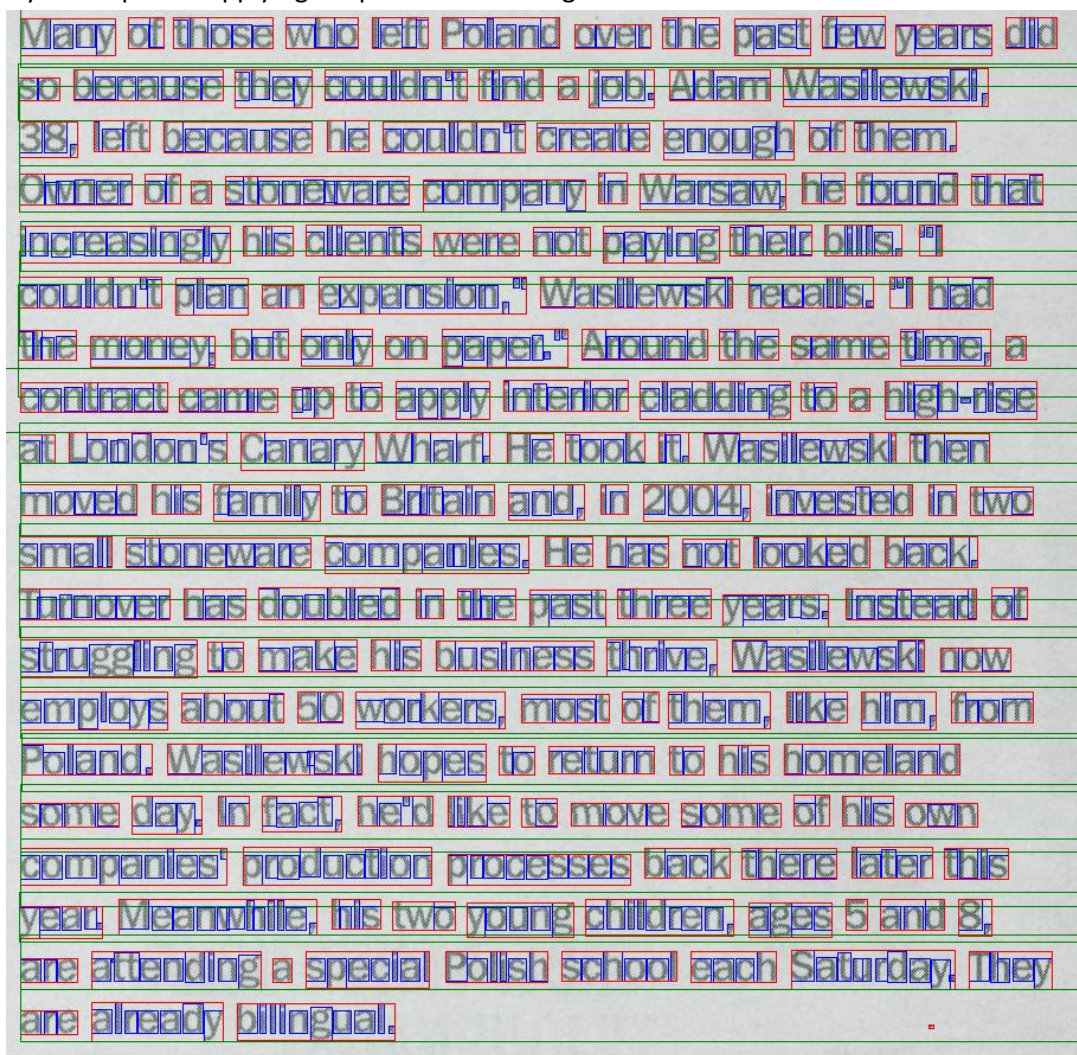
c. Pre-processing

Tesseract binaries image using Otsu thresholding method; in some cases, it is not robustness, below are results of AnalyseLayout step:



Many of those who left Poland over the past few years did so because they couldn't find a job. Adam Wasilewski, 38, left because he couldn't create enough of them. Owner of a stoneware company in Warsaw, he found that increasingly his clients were not paying their bills. "I couldn't plan an expansion," Wasilewski recalls. "I had the money, but only on paper." Around the same time, a contract came up to apply interior cladding to a high-rise at London's Canary Wharf. He took it. Wasilewski then moved his family to Britain and, in 2004, invested in two small stoneware companies. He has not looked back. Turnover has doubled in the past three years. Instead of struggling to make his business thrive, Wasilewski now employs about 60 workers, most of them, like him, from Poland. Wasilewski hopes to return to his homeland some day. In fact, he'd like to move some of his own companies' production processes back there later this year. Meanwhile, his two young children, ages 5 and 8, are attending a special Polish school each Saturday. They are already bilingual.

IPoVnOCRer uses Adaptive Thresholding method instead of Otsu, and below are results of AnalyseLayout step after applying adaptive thresholding.



Many of those who left Poland over the past few years did so because they couldn't find a job. Adam Wasilewski, 38, left because he couldn't create enough of them. Owner of a stoneware company in Warsaw, he found that increasingly his clients were not paying their bills. "I couldn't plan an expansion," Wasilewski recalls. "I had the money, but only on paper." Around the same time, a contract came up to apply interior cladding to a high-rise at London's Canary Wharf. He took it. Wasilewski then moved his family to Britain and, in 2004, invested in two small stoneware companies. He has not looked back. Turnover has doubled in the past three years. Instead of struggling to make his business thrive, Wasilewski now employs about 50 workers, most of them, like him, from Poland. Wasilewski hopes to return to his homeland some day. In fact, he'd like to move some of his own companies' production processes back there later this year. Meanwhile, his two young children, ages 5 and 8, are attending a special Polish school each Saturday. They are already bilingual.

d. How to improve in future?

Currently, tesseract cannot segment paragraph/column successfully (in my test ☺, see results below). In my opinion, we can own this step by using rule-base (abnormal-distance between two words on each line ...) to detect tab-stop (named by Ray Smith) in which separated two-columns in document.

Later in a pub, Izabela Chudzicka, the former bank clerk, talks about the fund raiser, which she's helped to organize in the past. In an accent betraying a hint of an Irish brogue, she says the Poles and the Irish have a lot in common. "We socialize in the same way. We know how to talk to each other. We understand each other's sense of humor. It's great!"

Listen to the rhetoric of politicians across Europe and you won't hear the relationship between Poles and their host countries described in such friendly terms. In 2005, Philippe de Villiers, leader of France's Euro-skeptic Mouvement pour la France, darkly warned of the "Polish plumber and Estonian architect" triggering "the demolition of France's social and economic model." Before the E.U. admitted 10 new members back in 2004, populist fears of unwashed hordes stealing jobs from local workers led most of the old E.U. countries, including Germany, Austria and France, to keep their labor markets closed.

Citizens and media in those countries warned that an influx of workers from Eastern Europe would undermine local economies, steal jobs and bankrupt the welfare system; the impact has been quite different. Polish migrants like Chudzicka have integrated seamlessly; 75%, in one survey, said the Irish have "made them feel welcome."

Well they might, for there is no evidence that incoming workers have systematically displaced locals or stolen their jobs. Unemployment is higher in France, from which Poles were turned away, than in Britain where they were welcomed. The jobless rate in Ireland is just 4.5%, and job vacancy rates reported by Irish businesses in the past two years have actually risen, from 11% to 17%. The positions migrants are filling, economists say, are either ones that locals don't want, or new positions altogether. In fact, the infusion of educated labor drove growth in host countries' most dynamic sectors.

B. Sample (refer to solution)

a. Sample1

i. Recognize

ii. Analyze Layout

b. Sample2

i. Recognize

ii. Analyze Layout

c. Sample3

i. Recognize