

Ultra Accurate Business Card OCR

About Us:

FullContact is solving the world's contact information problem. We're a 2011 TechStars Boulder graduate that provides APIs and applications that allow businesses and consumers to keep their contact information up to date automatically. We're a team of 45 people located in LoDo across the street (approximately 23 paces) from the Wynkoop Brewery.

About You:

You enjoy solving hard problems and figuring out how to make things work. You have a rough understanding of NLP and Machine Learning concepts, or you're excited to learn them in a fast-paced environment.

Background:

FullContact has a stated mission to become the premier contact management system of record for individuals and businesses. This is done by keeping all their contact information in one place and automatically up-to-date and making this contact information available everywhere they need it.

FullContact has developed a human based business card transcription service that processes thousands of business cards per day, but this takes time and is costly. We want to augment our human transcription with an OCR system to make the whole system faster and more affordable.

Project:

Deliver a solution that improves upon standard open source OCR libraries to create an ultra accurate scanning solution. The project will include image manipulation: scoring of image sharpness, rotation of image, image cropping, and color manipulation for optimal scan.

The project will also include use of open source OCR libraries such as Tesseract OCR for character recognition. Use of NLP and Machine Learning for identifying and classifying data elements on a card will be used.

Solution:

The ideal solution will be able to:

- identify card images that are too blurry for OCR or human analysis,
- optimize the card image for OCR analysis,
- identify and extract text from the image and properly categorize (i.e. email, phone, first name, last name, company, etc.),
- identify likely language of the characters on the card (English, Russian, Chinese, etc.),
- identify the likely location (country) that the card owner is located.

Location:

LoDo (across the street from the Wynkoop near Union Station)

Summer Internship:

A limited number of paid summer internships will be available upon course completion.