

Photo Deduplication

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 25 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 enjoy working with and building RESTful APIs.

Background:

FullContact has a stated mission of *Solving the World's Contact Information Problem*. We have a Cloud Address Book in which users sync their contacts with Gmail, CRMs, iCloud, and the like. Users also share contacts with each other. A contact can have many photos, and we're looking for a high-throughput, low-latency means of identifying "duplicate photos", where a duplicate photo is considered visually identical with another.

Solution:

This isn't simple. We can't simply go all $O(n^2)$ on it and expect to delight our users. We want to use Hadoop, Storm or some combination of the two to do this in real-time, using computed photo similarity metrics like Perceptual Hashing and unsupervised learning algorithms to cluster photos together.

You will work with stakeholders to design a system that interfaces with our Cloud Address Book API, Amazon S3, and several other technologies. The ideal solution will have high precision and recall while operating at high scale.

Location:

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

Summer Internship:

Project members are encouraged to continue the project through the course of the summer in the form of a paid internship.