

FullContact - NLP / Information Extraction

About Us

1. We're a startup located in LoDo (23 paces away from the Wynkoop Brewery).
2. We're [TechStars graduates](#).
3. [We're solving the world's contact information problem](#).
4. We <3 Big Data.
5. We <3 machine learning.
6. We were founded in 2010.
7. We're lean, mean and venture-funded.
8. We're not afraid of bleeding-edge technologies

About You

1. You want to get experience with machine learning, natural language processing, and data mining
2. You're familiar with the JVM.
3. You know Java, but you prefer other JVM languages like Groovy, Clojure, Scala, JRuby, or Jython.
4. You can't think of a better version control system than Git.
5. You took more math than you needed to graduate.
6. You find code without good test coverage to be brittle and unpleasant.
7. You know your way around a *nix shell like bash or zsh.

Problem Description

At FullContact, we're solving the world's contact information problem. This means we deal with contact information, and ~~lots of it~~ **terabytes of it**. We have to process structured data, semi-structured data, and completely-unstructured data like free text. This is where you come in.

We need your help to design and implement an information extraction system which can classify and extract meaningful contact information from free text. Examples of this are street addresses, company names, places, names, titles, phone numbers, and more. We have several ideas on how to do this (e.g. Markov Logic Networks, Viterbi Path analysis, grammar-based or statistical approaches) but we're looking forward to see what suggestions you may have. By the end of the project, we think you should be able to build a system which rates itself by [F-score](#), using a test set of manually-classified text.

We think this is an ideal project for a team of **2-3 people** that **love data mining**, but **we'll listen to you** if you think it should be more or fewer.

You should choose this project because...

1. You'll get to experience life at a venture-funded startup.
2. You will get to work with cutting-edge Big Data technologies like Hadoop and Storm.
3. You will get to work with cool non-relational data stores like Redis, HBase and MongoDB.
4. You'll get to use alternative JVM languages like Groovy and Clojure.
5. You'll gain practical experience applying information extraction and natural language processing ideas to real-world problems.
6. We're a fun bunch of people.
7. If all goes well, we might even hire you after the project!

Questions?

Email me at dan@fullcontact.com