Title - **Entity Inferencing using Statistical Techniques & Machine Learning**

About the Company

FullContact is the most powerful fully-connected contact management platform for professionals and enterprises who need to master their contacts and be awesome with people. FullContact's cross-platform suite of Apps and APIs enhance contacts with 360° insights, while keeping them organized, in-sync, up-to-date, and safe.

In 2010, FullContact launched its first contact management solution to 100 beta users and began a journey to solve the world's contact information problem. Since then, FullContact has become the identity resolution leader— managing billions of contact records and helping millions of professionals gain complete customer insight by turning incomplete data into a full person or company record. FullContact is dedicated to providing the first and only Human to Human platform, with a suite of customer intelligence APIs that allows individuals and organizations to create authentic relationships and be awesome with people.

We're a team of 200+ people located across the globe in Denver, Texas, Latvia, Tel Aviv and India, headquartered in LoDo (downtown Denver) down the street from Union Station.

Background

FullContact has built the leading engine for resolving identities of individuals and businesses. Our systems are capable of crunching through terabytes of data and resolve identities based on billions of observations. If we have more inferencing capabilities but around fuzzy matching and statistical methods, we could materially improve our identity resolution capability.
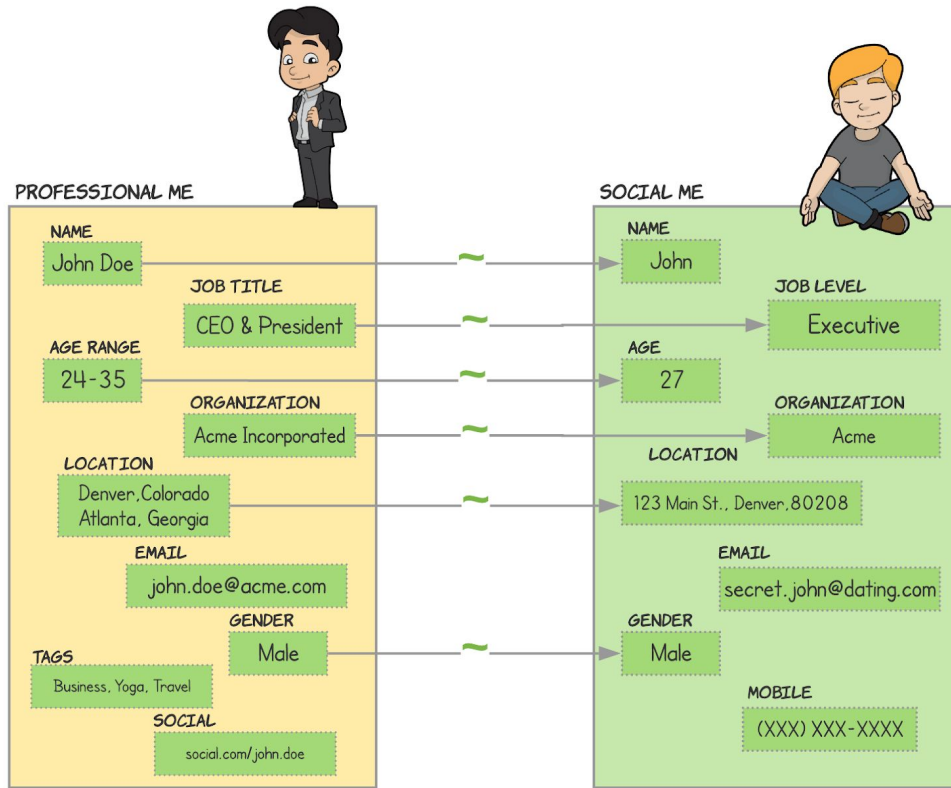
Task Description

Figure 1

In the above image, Yellow represents John Doe's professional profile and Green represents his personal profile. There are no explicit connections between the profiles, but with enough inferencing, we can predict that both profiles belong to the same person, at some level of confidence.

The task in hand is to match similar profiles based on fuzzy inferences, while maintaining precision and recall at an acceptable threshold. This would including building inferencers for field types like bio (by building topic models), location (by using geocoding techniques), job titles and organizations (by statistical learning) or names (using inferences based on hypocoristics). There are so many indicators, and some of them are strong and some weak.

The goal of the project is to create an inference system that matches two person profiles similar to one on the example shown above - it could be part of the whole inferencing system or inferences on a couple of fields which has got maximum bang to buck for identifying similar profiles.

Skills

1. Must haves
    a. Experience in computer programming in some form
    b. Interest in working with data
    c. A curious mind for exploration
    d. Strong propensity to scientific methods

2. Nice to haves
    a. Interest in machine learning and statistical modelling

Ideal Team size

3 - 5 working as a team

Location

FullContact
1755 Blake St #450
Denver, CO 80202
(2 blocks from Union Station)