Kyle Hubbard

Final Project Proposal

April 2, 2015

I will be working on a job title classification problem. One of the products that I am developing at work is a compensation tool for our clients to use when benchmarking employee pay. The job title data we have is not uniform thus one part of the project will be clustering similar job titles making them conform to a known list of job titles. This list will be downloaded from a third party either the [BLS](http://www.bls.gov/soc/), [ERI](http://www.erieri.com/SalaryAssessor#positionsincluded) or one of the other [job sites](http://developer.careerbuilder.com/) that have an API with job title information. Once the job titles have been made uniform I would like to classify them into a standard hierarchy or company organization chart. The US Department of labor and BLS publish job title data and hierarchies through [O\*NET](http://www.onetcenter.org/taxonomy.html). I will be using this as the foundation for my hierarchy. There are several challenges to overcome; one challenge is how to appropriately classify emerging job titles that are not currently covered by the O\*NET. An additional challenge is to appropriately cluster similar job titles while excluding unrelated job titles. For example CEO and Executive Assistant to the CEO are prone to being clustered together when they should not be. The final objective will be to develop a framework to appropriately classify new client job titles into the hierarchy to make full use of the available data in building a robust compensation tool.