Improved Linkedin Search

**Abstract:** I have seen multiple flaws in linkedin search including location, organisation and most importantly compound keywords, there is a significant lack of machine learning in search results based on mixed keywords.

Moreover there are certain limitations of search to registered non-premium users, such as limited results, and low visibility people with low profile ratings.

I am trying to implement a linkedin search which will likely return more statiscally signifant results based on user keywords through different methods.

**Data:** Data is collected from google custom search application module which is fetched to data frame through JSON based on the user input keywords.

The data frame is then cleaned by removing redundant data, unformatted data and then by arranging it in priority order as defined by user on input which is either based on name, location, organisation, or a keyword defined by user.

#R Code:

#Fetching data from Google Custom Search: Sample Document 1: 10 Results : Page 1

#User Input: Keyword: Alice New york C++

document <- fromJSON(txt='<https://www.googleapis.com/customsearch/v1element?key=AIzaSyCVAXiUzRYsML1Pv6RwSG1gunmMikTzQqY&cx=009462381166450434430:ecyvn9zudgu&q=adamnewyorkc>++')

#Data Frame Document contains results table

result <- document$results #Results stored in data frame

#Data Cleaning

#Removing redundant and useless variables

result$GsearchResultClass = NULL

result$content = NULL

result$title = NULL

result$clicktrackUrl = NULL

result$formattedUrl = NULL

result$unescapedUrl = NULL

result$richSnippet$metatags = NULL

result$richSnippet$hcard$photo = NULL

result$visibleUrl = NULL

result$titleNoFormatting = NULL

result$Organisation$hcard$title = NULL

#Reformatting by naming

names(result$contentNoFormatting) <- "Content"

names(result$Organisation$person)[1] <- "Organisation"

names(result$Organisation$person)[2] <- "Location"

names(result$Organisation$person)[3] <- "Role"

names(result$Organisation$hcard)[1] <- "Name"

View(result)

**Procedure:** After the data is obtained and cleaned the next part is extracting useful insights from data. We are going to use text mining to extract important keyowords and then apply modeling through various techniques.