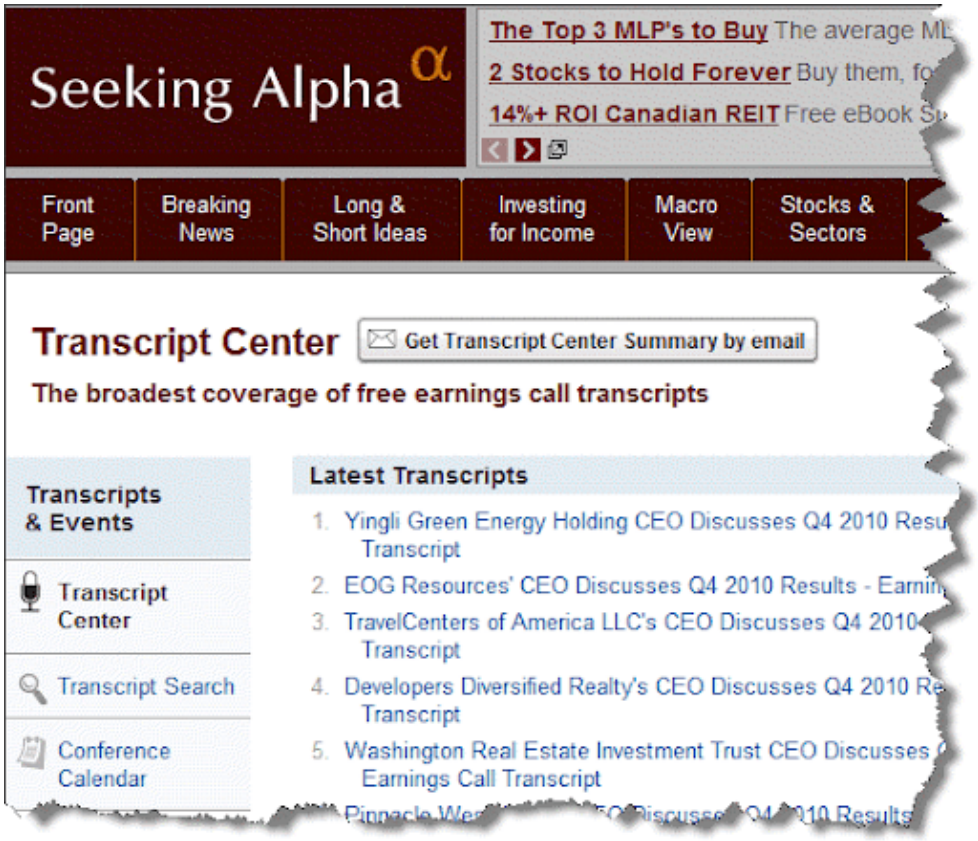
MSA 8040: Final Project



Background:

Recently, text mining has attracted lots of attentions in both academia and industry, especially in FinTech area. There are many algorithms which have been developed to either correctly predict company returns or to create reasonable investment index. Conference earning call scripts can provides lots of textual information for decision maker. It consists of both presentations and Question-Answering parts. Presentations is usually made by the executive staff from a company (usually CEOs); in Q&A part, participants like investors who will ask lots of questions and the company participants answering their questions.

This final project is a group-based (group size at most 5) project. The task is to build an entity relational database for the conference earning call script, named CECS. You should first design your own ERD, using the normalization techniques. So, each entity in your entity relational database should be at least in 3-NF. Your database should contain at least the following attributes:

1. Ticker: ticker name of the company
2. Company, the detailed name of the company
3. Title, the title of the conference call script

4. Date, the date when the conference call script takes place
5. Time, the specific timestamp when the conference call script starts
6. Section: presentation or Q&A
7. Speech: textual information (e.g., content)
8. Participant name: the name of the participants
9. Participant type: CEO, Company Staff, Analyst, or Others
10. Participant Organization: the name of the participant organization

You should decompose the above attributes (not limited to these 10 attributes) into different entities and each entity contains reasonable attributes. After your database schedule is completed, you should be able to scrape data from https://seekingalpha.com/earnings/earnings-call-transcripts for year 2019 and 2020. Once your CECS database is completed, you should be able to write SQL query to answer questions similar to the following:

1. How many conference calls happen on the Q1, 2020.
2. Given a ticker name, e.g., FUV, how many conference calls are in 2020.
3. Given a ticker name and date, e.g., FUV, how participants in the conference call and who are them? And further display their speech, given the name of a participant.
4. Etc.

Requirements:

1. iCollege submission deadline: 12/10/2020
2. Presentation date: lecture time @ 12/10/2020 (including your data schema, ERD and demo for several queries)
3. iCollege submission files (group-based, only one need to submit): ERD and code for web scraping and database construction.
4. Peer evaluation form (individual-based, each member should submit their own evaluation for the other teammates)

Peer-Evaluation (this is for the final group project):
To better achieve fairness in the class, at the end of the course you will be asked to evaluate yourself and the other members of your group on completing the project. These ratings are used for gauging team members' contributions. The grade you and your group members receive will depend in part on these peer evaluations. Rate each member based on the following criteria: (1) participation in group activities, (2) quality of work, (3) quantity of work, (4) finishing assigned work on time, and (5) ability to work as a team member. Please use the following scale to assign scores:

5	Exceptional effort, above and beyond the call of duty
4	Above average effort
3	Normal effort (this is the expected score!)
2	Below average effort
1	Unacceptable effort

Then, submit the following note to the instructor:

Your Name: _____ Score: _____
Team Member #2: _____ Score: _____
Team Member #3: _____ Score: _____
Team Member #4: _____ Score: _____

Team Member #5: _____ Score: _____

Note: Please include a brief reason for any group member scoring either a "1" or a "5." I expect everyone to be thoughtful and diligent in completing this evaluation. ***You may get ZERO for the project if you receive "1"s from all other group members.***