Homework 6

1. From the Indiegogo (https://webrobots.io/indiegogo-dataset/) dataset you need to download at least 5 JSON files. Use the content of "tagline" or "title" from JSON files.

2. Extract the article title from your downloaded dataset and use "bag of words" to convert the article title into set of words.

3. Use LSA or LDA to cluster them. You need to do research online about how using LSA and LDA and do they work, we did not develop them in the class, to enforce you do the research on your own.

4. Visualize the article clustering results, in a dendrogram or heatmap. Please be sure to show a clean dendrogram and not a sloppy dendrogram. Delivering an imbalanced unclear dendrogram, or a heatmap with too small text, will result in reducing your grade. You might need to annotate your dendrogram by hand and add information to it.
Besides, please spend time and create a proper visualization for your experiment. At this time, you should be able to create proper visualization and do not rely on default font settings.

You need to prepare a report on your tasks and findings along with a video file describing what you have done. You can copy paste your codes, its results and your description into a Word document, Python Notebook or you can use R notebook.

Your deadline for delivering this home work is written on the blackboard online. Please send your questions to TA and if required to RA.