
ASSIGNMENTS

GOOGLE COLABORATORY

- ▶ Use Colaboratory (<https://colab.research.google.com/notebooks/welcome.ipynb>) for all your programming assignments
- ▶ Use Python 3 notebook
- ▶ Keep the corpus that you may use for the assignment in a location that can be accessed by Colaboratory. Since the corpus is the same for everyone, you may choose to use a single location
- ▶ Once you complete your assignment, share the link of the notebook with me (ramaseshanr@gmail.com). The date of sharing the notebook with me will be considered as the submission date and time
- ▶ If you have any questions on colaboratory, read the FAQ :)
- ▶ Submitting the code through MOODLE is NOT accepted
- ▶ Do NOT share your code through email

CONSTRUCT THE BINARY INCIDENCE MATRIX USING THE DOMAIN KNOWLEDGE

The corpus contains the questions sourced from the Internet. The corpus (271 text files) contains questions from Kinematics class of physics problems.

In this assignment, you need to develop a python program that uses the knowledge related to Kinematics and build a table similar to the one shown on the right for all the documents in the corpus.

The program should be able to read each problem, capture the known values (such as speed=10m/s, time=5s) and fill the respective cells in the table. For example, if you find 10 m/s for document 1, fill the speed with value row for D1 as 1.

Please note that problems may or may not contain all nine terms listed. The corpus is available at the *NLP 2019 Moodle* page. The corpus may contain duplicate entries.

You may use any NLTK or any equivalent for this assignment

Terms	D1	D2	D3	D4	D5
Speed with value	1	0	0	0	0
Distance with value	0	1	1	1	1
Acceleration with value	0	0	0	0	0
Time with value	0	1	0	1	0
Speed in question	0	1	1	1	1
Time in question	1	0	1	0	0
Distance in question	1	0	0	0	0
Acceleration in question	0	0	0	0	0