## Different Models For Multi Documents summarization.

* potara -> Working Fine for less documents , but summaries are't good for them.They have mentioned it works great for

more documents . For more documents we faced an error related to parallel computing .

The model is good as it takes into account parallel implementation at the backside.

Also a lot of errors were faced in making it run on windows.Suggested ti use uubuntu for this.Errors ayenge bt kum

[ILP based]

https://github.com/sildar/potara

=========================================

[2021]

* mgSum : didnt have all the tar,val .sent files which it required. Might be in the 2000-300 file.

Also a lot of training . reqd .Tried training .

GPU's will be required ,and we don't have pretrained labels for them.

If choosing and some how making this model work , we will nedd to train it rigourosuly over the time from now to July

https://github.com/zhongxia96/MGSum

====================

* ALLEN -AI primera

https://github.com/allenai/primer

The most primising one yet.

Paper availabe ,just google it and do read if pssbl

Fine tuning this model not possible or very difficult , so even it works we can just train it more to uit our

requirement.

USE UBUNTU FOR THIS.

Error observed -> assertion error in some internal files.

NOTE : LESS/No support is availble for all above models.Error handling mein time jaata hai.

=======================

https://github.com/anushree1808/Abstractive-Multi-Document-Text-Summarization

* ANushree :Local college project of someone.

We found this in the morning 2 am of our presentation , it doesn't have many forks and ReadME but yet to give this a try.

======

* OWN MODEL DEVELOPMENT

We tried to make our own model using smaller parts from other repos ,

The approach is as follows:

-We have documents ready.

-step 1 : Find best document

A lot of pre data processing will be required.Figure it out

-step 2 : Make sentence clusters .(Towards DataScience)

- step 3: from sentence clusters generate an ILP problem which will give us the multi-doc summary.

===

Step 3:(Most important)

https://github.com/siddBanPsu/AbTextSumm

We were stuck with the language modelling which was required for the input.(KENLM) i.e. the input text must be in KENLM binary format.

for KENLM: (https://kheafield.com/code/kenlm/) ,, https://github.com/kpot/kenlp

[ILP based ]

============================================

* https://github.com/Oceandam/EMSum

===============

* https://github.com/boudinfl/takahe

Above is a multi-sentence compression .

Idea was to take text of all doucments into one dcument and then process it using above model.

====

Based on the above idea

* https://github.com/johnsonj561/K-Means-Document-Summarization
* K-means clustering :: it summarizes sentences too .

(An extractive one) To uderstand how extractive differs from abstractive.

==================

* Based on neural net ::

https://github.com/ayushoriginal/Multi-Document-Summarization

Tough one to understand . ANd also training ka issue rahega.GPUS will be reqd.