

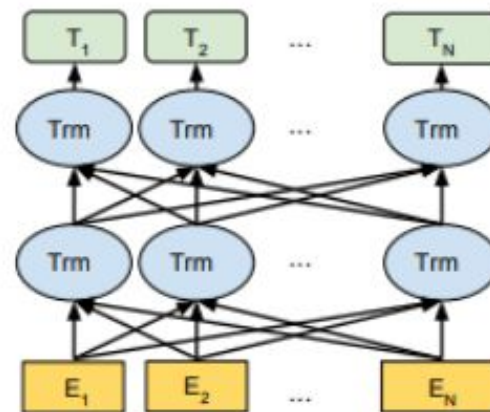
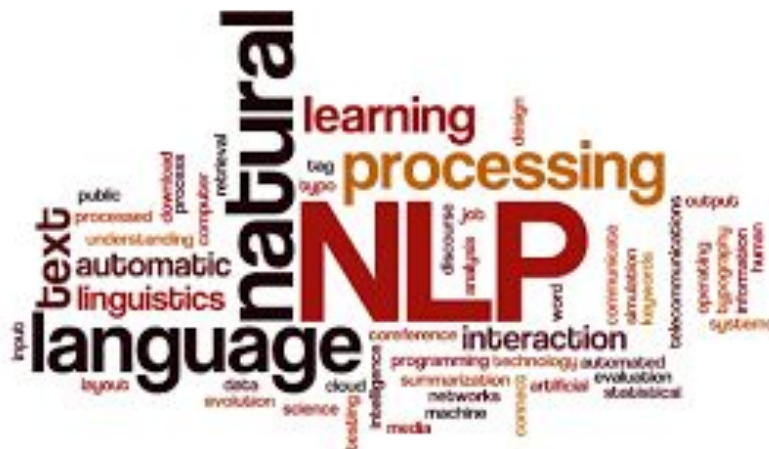
Text Summarization From Movie Dialogues

...

Data-X 2 - Plaksha TLF

Akbar Surani
Anmol Dureha
Pradeep Kumar
Rakesh Siri

Inspiration



BERT Architecture

Problem Statement / Approach

**Text Extraction Based Summarization of Movies Using Text
Ranking and Clustering on Sentence Embeddings (BERT)**

Summarization - Extractive vs Abstractive

Extractive - Finding most informative sentences from a body of text, which can be used to generate summary

Abstractive - Generating phrases that are semantically consistent (The way humans summarize)

Use Cases

Rhetorica ad Herennium put it 2,100 years ago: 'We wish to have our hearer receptive, well-disposed, and attentive (*docilem, benivolum, attentum*).’ To be civilised was to speak persuasively about the things that mattered: law and custom, loyalty and justice.

This vision of puppeteers effortlessly pulling everyone else's strings — however much it might fulfil both geek fantasies and Luddite nightmares — is distinctly dubious

Underpinning this was neither honour nor idealism, but pragmatism embodied in a five-part process. Come up with a compelling proposition, arrange its elements in elegant sequence, polish your style, commit the result to memory or media, then pitch your delivery for maximum impact. Short of

Auto-select the most informative or relevant quotes for online articles



Generating quotes/dialogues for fanfiction

VINCENT
You'll dig it the most. But you know what the funniest thing about Europe is?

JULES
What?

VINCENT
It's the little differences. A lotta the same shit we got here, they got there, but there they're a little different.

JULES
Examples?

VINCENT
Well, in Amsterdam, you can buy beer in a movie theatre. And I don't mean in a paper cup either. They give you a glass of beer, like in a bar. In Paris, you can buy beer at MacDonald's. Also, you know what they call a Quarter Pounder with Cheese in Paris?

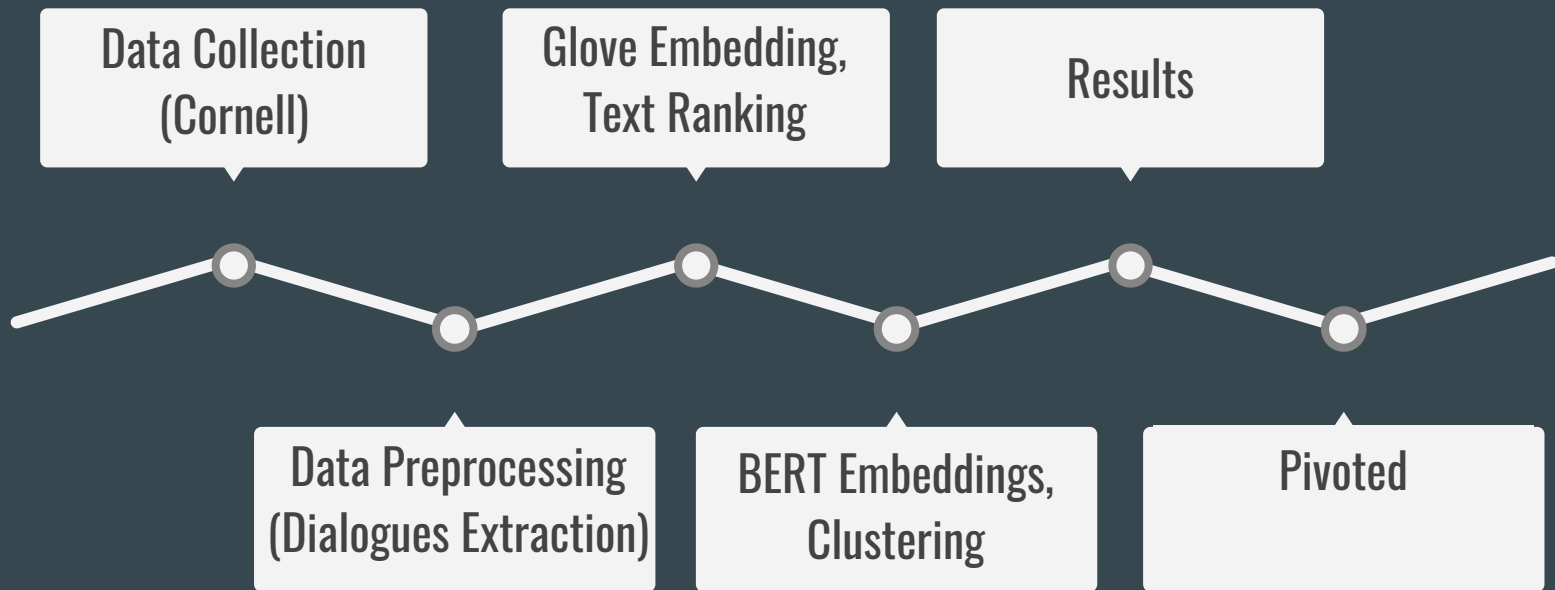
JULES
They don't call it a Quarter Pounder with Cheese?

Suggesting characters for movie script transcription



Suggesting dialogues to include in Movie Trailers

Our Journey



Dataset - Cornell Movie Dialogues Corpus

The room is large and humid. Rows of salvaged junk are stacked neatly against the walls. Two large fans whirl above their heads.

LEON

Okay if I talk?

Holden doesn't answer. He's centering Leon's eye on the machine.

LEON

I kinda get nervous when I take tests.

HOLDEN

Don't move.

LEON

Sorry.

He tries not to move but finally his lips can't help a sheepish smile.

LEON

Already had I.Q. test this year -- but I don't think I never had a...

HOLDEN

(cutting in)

Reaction time is a factor in this, so please pay attention. Answer quickly as you can.

Leon compresses his lips and nods his big head eagerly. Holden's voice is cold, geared to intimidate and evoke response.

HOLDEN

972 Movies

14035 Characters

304713 Utterances

Data Processing - Extracting Dialogues

The room is large and humid. Rows of salvaged junk are stacked neatly against the walls. Two large fans whir above their heads.

LEON

Okay if I talk?

Holden doesn't answer. He's centering Leon's eye on the machine.

LEON

I kinda get nervous when I take tests.

HOLDEN

Don't move.

LEON

Sorry.

He tries not to move but finally his lips can't help a sheepish smile.

LEON

Already had I.Q. test this year -- but I don't think I never had a...

HOLDEN

(cutting in)

Reaction time is a factor in this, so please pay attention. Answer quickly as you can.

Leon compresses his lips and nods his big head eagerly. Holden's voice is cold, geared to intimidate and evoke response.

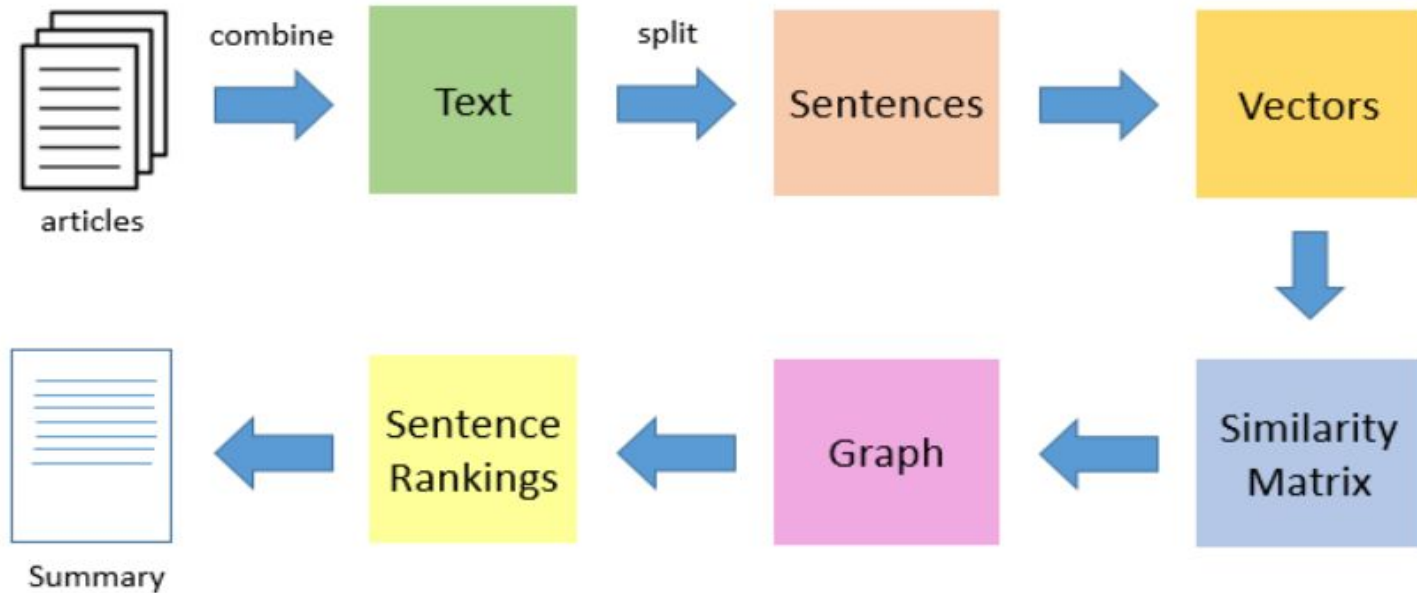
HOLDEN

[22]: ['Screenplay by',
'1420 No. Beachwood Drive Hollywood, Calif. 90028 *****',
'Okay if I talk?',
'I kinda get nervous when I take tests.',
'Don't move.',
'Sorry.',
'Already had I.Q. test this year -- but I don't think I never had a...',
'(cutting in) Reaction time is a factor in this, so please pay attention. Answer quickly as you can.',
'You're in a desert, walking along in the sand when all of a sudden you look down and see a...',
'What one?',
'What?',
'What desert?',
'Doesn't make any difference what desert -- it's completely hypothetical.',
'But how come I'd be there?',
'Maybe you're fed up, maybe you want to be by yourself -- who knows. So you look down and see a tortoise. It's crawling towards you...',
'A tortoise. What's that?',
'Know what a turtle is?',
'Of course.',
'Same thing.',
'I never seen a turtle.',
'But I understand what you mean.',
'You reach down and flip the tortoise over on its back, Leon.',
'You make these questions, Mr. Holden, or they write 'em down for you?',
'The tortoise lays on its back, its belly baking in the hot sun, beating its legs trying to turn itself over. But it can't. Not without your help. But you're not helping.',
'Whatcha mean, I'm not helping?',
'I mean you're not helping! Why is that, Leon?',
'(V.O.) It was 97 degrees in the city and no hope of improvement. Not bad if you're a lizard. But two hours earlier I was drinking Acquavit with an Eskimo lady in North East Alaska. That's a tough change to make. It was so good, I didn't want to leave, so I left a day early.',
'Contact. This is Blade Runner One calling Com-fast 27.',
'Blade Runner One, stand by please.',
'Deckard.',
'Yah, Gaff.',
'Where the hell you been?',
'You know where I been. I been on vacation.',
'Next time you go on vacation, do me a favor, let us know where it is.',
'What's up?',
'Holden got hit.',
'Bad?',
'Severed spine. You'd better get in here. Bryant's waiting for you.',
'I'll see you in a minute.',
'(V.O.) I-X-4-P-D referred to as a Nexus-6, The Tyrell Corporation's new pride and joy. Holden was administering the Voight-Kampff test when one nailed him.',
'(V.O.) The Nexus-6 must be fast because Holden was as quick as they come. The report said there were six of them. Three males and three female. Led by a combat model called Roy Batty.',
'(V.O.) They escaped from the colonies two weeks ago. Killed twenty-three people and jumped a shuttle. An aerial patrol fou

Methods Used

- We used below mentioned methods to do unsupervised learning:
 - Glove Embedding
 - Text Ranking
 - BERT Embeddings

Text ranking using GloVe Embedding



Bert Embedding

- Pre-trained BERT-based text summarization model
- First embedding the sentences
- Running a clustering algorithm
- Finding the sentences that are closest to the cluster's centroids

Output

Movie - A Few Good Men

Good, Jo, that's confidence inspiring." "Thank you, sir." Someone who can really look into this. (beat) Myself." "Certainly, sir." She disposed of three cases in two years. She's not cut out for litigation. All passion, no street smarts. "Nothing to be sorry about, Sherby. "(calling back) I don't think so, Danny." I will get it. "Guantanamo Bay, Cuba. Commander Joanne Galloway. She's the one who had 'em brought up here. I'm lead counsel. This is Sam Weinberg. Going on nine months now. (beat) I see. "Commander, Lt. They'll need a lawyer. You with me?" Twelve years. "My job is to make sure you do your job. Read the letters. "Sorry, I always forget that." "(V.O.) Dear Sir,"

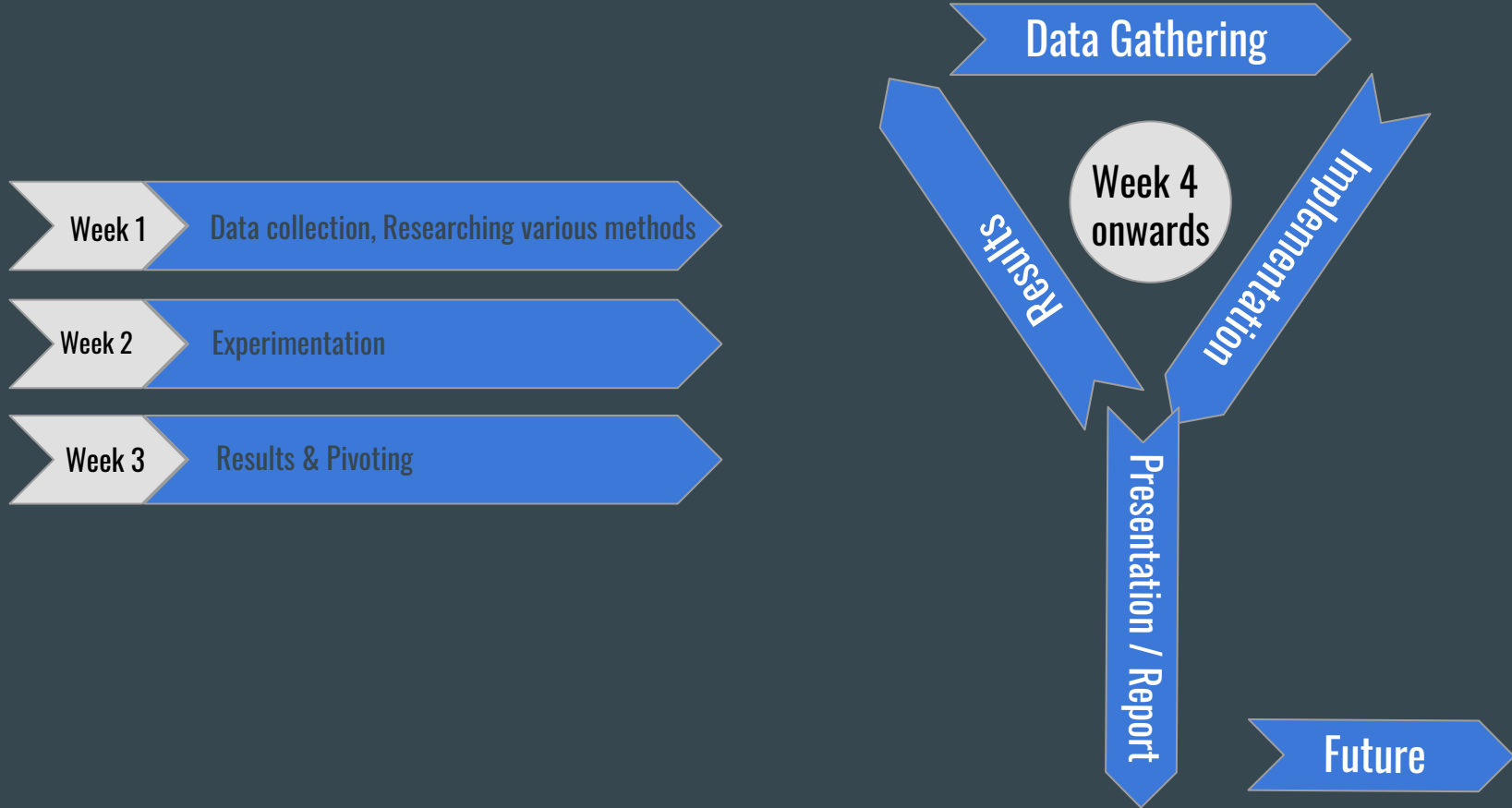
About being marines. Who the fuck is William T. Santiago. "You're appalled? I've got a better idea. Let's -- on second thought -- Windward. (calling out) Tom!" "Wait a minute, Tom." "(continuing) Don't call the President just yet. "(continuing) Matthew, sit, please." "(continuing) Matthew, siddown. "No, thank you, I can't throw and catch things." Dawson and Downey. I don't think your clients murdered anybody. (continuing) What a pity. ": M.P. Officer on deck, ten-hut." "Sir, Lance Corporal Harold W. Dawson, sir. (to

) Is this your signature? Harold, what's a Code Red?" What's a garden variety Code Red? Was the attack on Santiago a Code Red? When all of a sudden...?) Did anyone see you call the ambulance? Were you there when the ambulance got there? "He went outside his unit, sir. "To train him, sir." Train him to think of his unit before himself. Dan Kaffee. Sailin' Jack Ross. I'll talk to you when I get back. "You gotta play 'em as they lay, Luther." "When Nancy gets back, you're my witness. The girl sat here, pointed, and said ""Pa"". She said ""Pa"". "You don't believe their story, do you? I don't even know who he is. Lieutenants Kaffee and Weinberg? (shouting) Yeah. We gotta take a boat?! "(shouting) No. No problem. The

Pivoting



Roadmap



Dataset - South Park Cartoon Dialogues



	Season	Episode	Character	Line
0	10	1	Stan	You guys, you guys! Chef is going away. \n
1	10	1	Kyle	Going away? For how long?\n
2	10	1	Stan	Forever.\n
3	10	1	Chef	I'm sorry boys.\n
4	10	1	Stan	Chef said he's been bored, so he joining a gro...

18 Seasons

70,000 + Utterances

Well Structured, Clean

Allows Analyses based on Characters

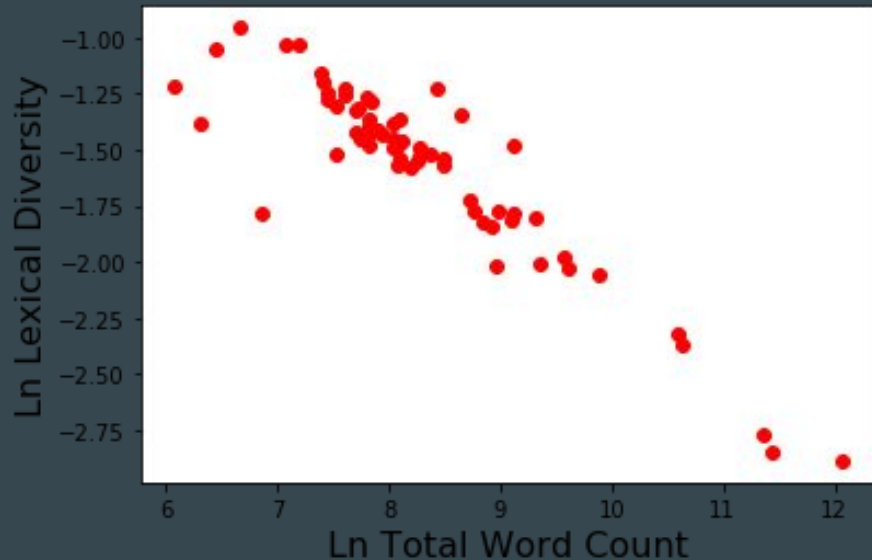


Problem Statement / Approach

Predicting the Speaker (Dialogue -> Speaker)

Quote Generation (Speaker -> Dialogue)

Exploratory Data Analysis



Log Lexical Density vs Log Word Count

Character		
Cartman		9774
Stan		7680
Kyle		7099
Butters		2602
Randy		2467
...		
Man 25	1	
Man 26	1	
Man 27	1	
Man 28	1	
A Banana	1	

Dialogue Frequency of Characters

Predicting The Speaker

- Methods Used:
 - Random Forest Classifier
 - Logistic Regression
- Input: Dialogue
- Output: Speaker

Quote Generator

- Methods Used:
 - n-grams
- Input: Speaker
- Output: Dialogue

Sample Output / Result

```
[ ] Generate_quote(kyle_bigram, 2, 'i', 12)
```



```
"i ca n't possibly have a kid put those babies ."
```

Sample Output / Results

```
[ ] Generate_quote(Cartman_trigram, 3, 'i', 12)
```



"i 'm not funny ! i should be fine . just , i did n't you tell anyone ! i can we , dog , i was gon na die



- "i do n't wan na stop feeling like i 'm looking for the olives ."
- "i 'll kick you in rancho de burritos rojos , south park ."
- "i see where have i like to do is exactly the border ! because i 'm trying to have another abortion , but mo-o-om , what a ninja when you can either give it back to our freedom ."
- 'i know . listen everybody at war , you think he has... no .'
- "i do n't you ever since wendy psst . perhaps you doing with you and with all right , cupid me ! yes ."

Thank You