

Getting Started with Text Data

Sequence Modeling in Text: Text Classification

My experience so far
has been fantastic!



Sequence Modeling in Text: Search Engines

Is iphone
better than
android?

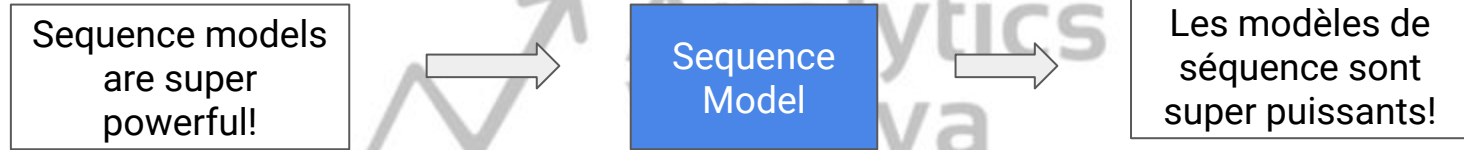


Sequence
Model



The screenshot shows a Google search interface. The search bar contains the text "Is iphone better than android?". Below the search bar, the top result is from "yourstory.com" with the title "iPhone Vs Android: Reasons to pick the iPhone over Android". The snippet text reads: "iPhone is faster than Android based phones. In tests, iPhone takes a lead once it starts to open apps that are stored in the background. And the faster iPhone 8 will be even faster. Apple's iPhone is leading because of better hardware and software integration and controlling both sides of the equation. Sep 14, 2017". Below this, there is a "People also ask" section with four questions: "Is Apple better than Android?", "Which is more user friendly iPhone or android?", "Do iPhones last longer than androids?", and "Is iPhone better than Samsung?". Each question has a downward arrow indicating it can be expanded. At the bottom, there are two more search results: one from "tomsguide.com" titled "10 Reasons the iPhone Beats Android | Tom's Guide" dated Apr 19, 2019, and another from "lifelwire.com" titled "iphone-vs-android-best-smartphone-2000309".

Sequence Modeling in Text: Machine Translation



Sequence Modeling in Text: Chatbots



Sequence Modeling in Text: Summarisation

A strong analyst has 6 main characteristics. One should master all 6 to be successful in the industry:

1.
2.
3.

Sequence
Model

6 characteristics of a
successful analyst

Examples of Text Data



Examples of Text Data

- Social Media Communications



Examples of Text Data

- Social Media Communications:

- Tweets
- Posts
- Comments

Anal
Vidh



Examples of Text Data

- Social Media Communications
- Conversations



Examples of Text Data

- Social Media Communications
- Conversations:
 - Messages
 - Emails
 - Chats



Examples of Text Data

- Social Media Communications
- Conversations
- Articles



Examples of Text Data

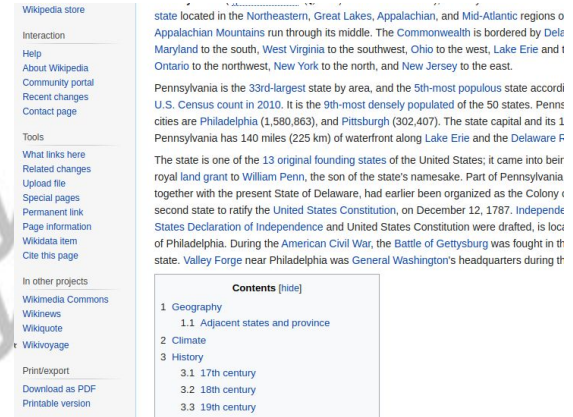
- Social Media Communications
- Conversations
- Articles:
 - News
 - Blogs
 - Transcripts



A screenshot of the Wikipedia page for the state of Pennsylvania. The page is in English and shows the standard Wikipedia layout. On the left, there is a sidebar with links to the Wikipedia store, interaction, help, and various tools. The main content area on the right contains the introductory text of the article, describing Pennsylvania's location, population, and history. A "Contents" table of contents is visible at the bottom of the main content area, listing sections like Geography, Climate, and History with sub-sections.

Examples of Text Data

- Social Media Communications
- Conversations
- Articles:
 - News
 - Blogs
 - Transcripts
- And many more!



The screenshot shows a Wikipedia page for Pennsylvania. The sidebar on the left contains links for 'Wikipedia store', 'Interaction', 'Help', 'About Wikipedia', 'Community portal', 'Recent changes', 'Contact page', 'Tools', 'What links here', 'Related changes', 'Upload file', 'Special pages', 'Permanent link', 'Page information', 'Wikidata item', 'Cite this page', 'In other projects', 'Wikimedia Commons', 'Wikinews', 'Wikiquote', 'Wikivoyage', 'Print/export', 'Download as PDF', and 'Printable version'. The main content area includes a paragraph about Pennsylvania's location, a paragraph about its size and population, and a paragraph about its history. A 'Contents [hide]' table of contents is visible at the bottom right of the main content area.

Contents [hide]	
1	Geography
1.1	Adjacent states and province
2	Climate
3	History
3.1	17th century
3.2	18th century
3.3	19th century

What does a Text data consist of?



What does a Text data consist of?

Anal
Vidhya



The image shows a screenshot of a news website and a Twitter feed. The news website has several headlines: "Rebels face down PM over Brexit trade deal", "Diver dies during rescue in Thai caves", "Police search for police officer in Salisbury", and "Merkel rival backs May over Brexit". The Twitter feed shows three tweets from Kanye West (@kanyewest) posted 9 hours ago. The first tweet says "This is the funnest car I've ever driven" with 691 replies, 3.3K retweets, and 34K likes. The second tweet says "I'm super chaaaaaarged. Bout to take this whole thing to mars" with 644 replies, 8.8K retweets, and 46K likes. The third tweet says "I heard these are really good for the environment 🙌" with 250 replies, 2.3K retweets, and 22K likes.

Rebels face down PM over Brexit trade deal
Seven cabinet ministers to oppose PM's plans
Theresa May faces the worst rebellion of her leadership today as rebel Brexiters attempt to force her to pull her Brexit deal from the House of Commons. The prime minister is planning to resign unless she can get the deal through. The rebels are demanding a second referendum on Brexit.

Diver dies during rescue in Thai caves
A Thai cave rescuer has died after being crushed by a rock while helping a group of soccer players escape from a cave in Thailand. The rescue operation has been ongoing for several days.

Police search for police officer in Salisbury
A police officer is being searched for in Salisbury after a report of a sighting. The officer is believed to have been involved in a protest in the city.

Merkel rival backs May over Brexit
A prominent German politician has backed Theresa May's Brexit deal, despite being a rival of the German Chancellor. The politician is a member of the CDU party.

KANYE WEST @kanyewest · 9h
This is the funnest car I've ever driven
691 3.3K 34K

KANYE WEST @kanyewest · 9h
I'm super chaaaaaarged. Bout to take this whole thing to mars
644 8.8K 46K

KANYE WEST @kanyewest · 9h
I heard these are really good for the environment 🙌
250 2.3K 22K

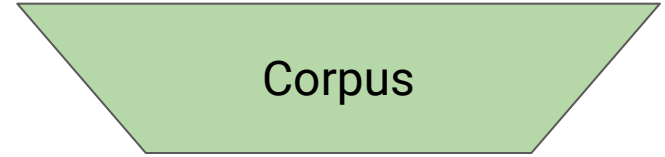
What does a Text data consist of?

- **Corpus:** A large set of texts

Anal
Vidhya



What does a Text data consist of?



- **Corpus:** A large set of texts



What does a Text data consist of?

- **Corpus:** A large set of texts
- **Document:** Single consolidated text file



Corpus

Document

What does a Text data consist of?

- **Corpus:** A large set of texts
- **Document:** Single consolidated text file
- **Paragraph:** Section of text in a file

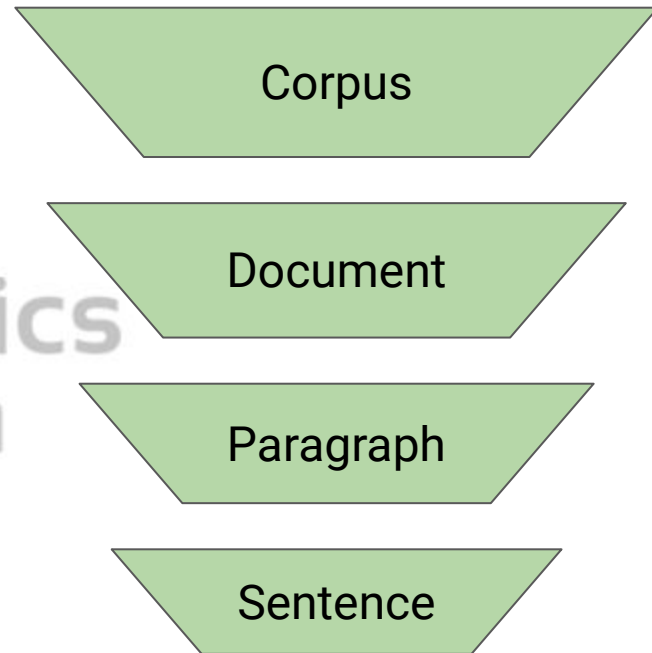
Corpus

Document

Paragraph

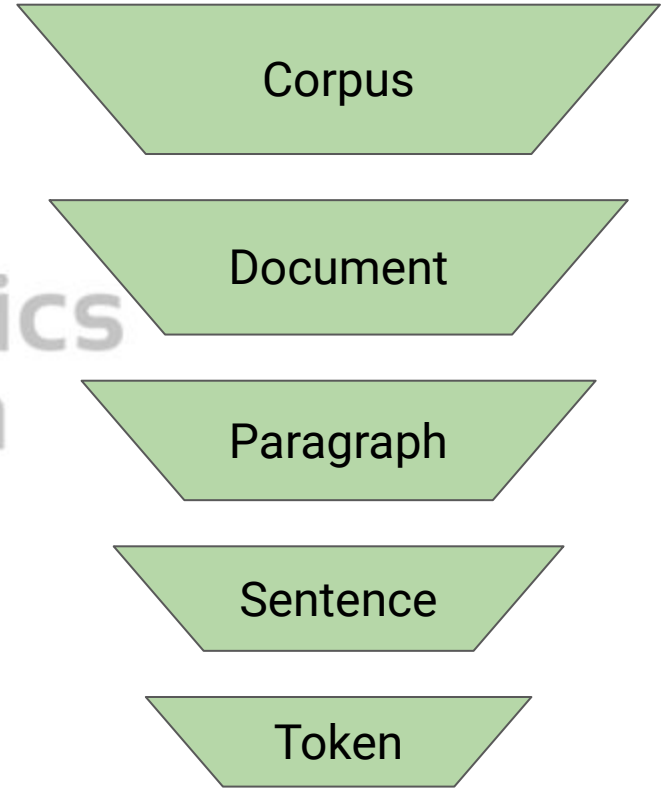
What does a Text data consist of?

- **Corpus:** A large set of texts
- **Document:** Single consolidated text file
- **Paragraph:** Section of text in a file
- **Sentence:** Group of words, complete in itself



What does a Text data consist of?

- **Corpus:** A large set of texts
- **Document:** Single consolidated text file
- **Paragraph:** Section of text in a file
- **Sentence:** Group of words, complete in itself
- **Token:** Smaller units of a text



 Thank You 