

Automatic Text Summarization

```
graph LR; A[What is Automatic Text Summarization] --> B[Methods]; B --> C[Evaluation];
```

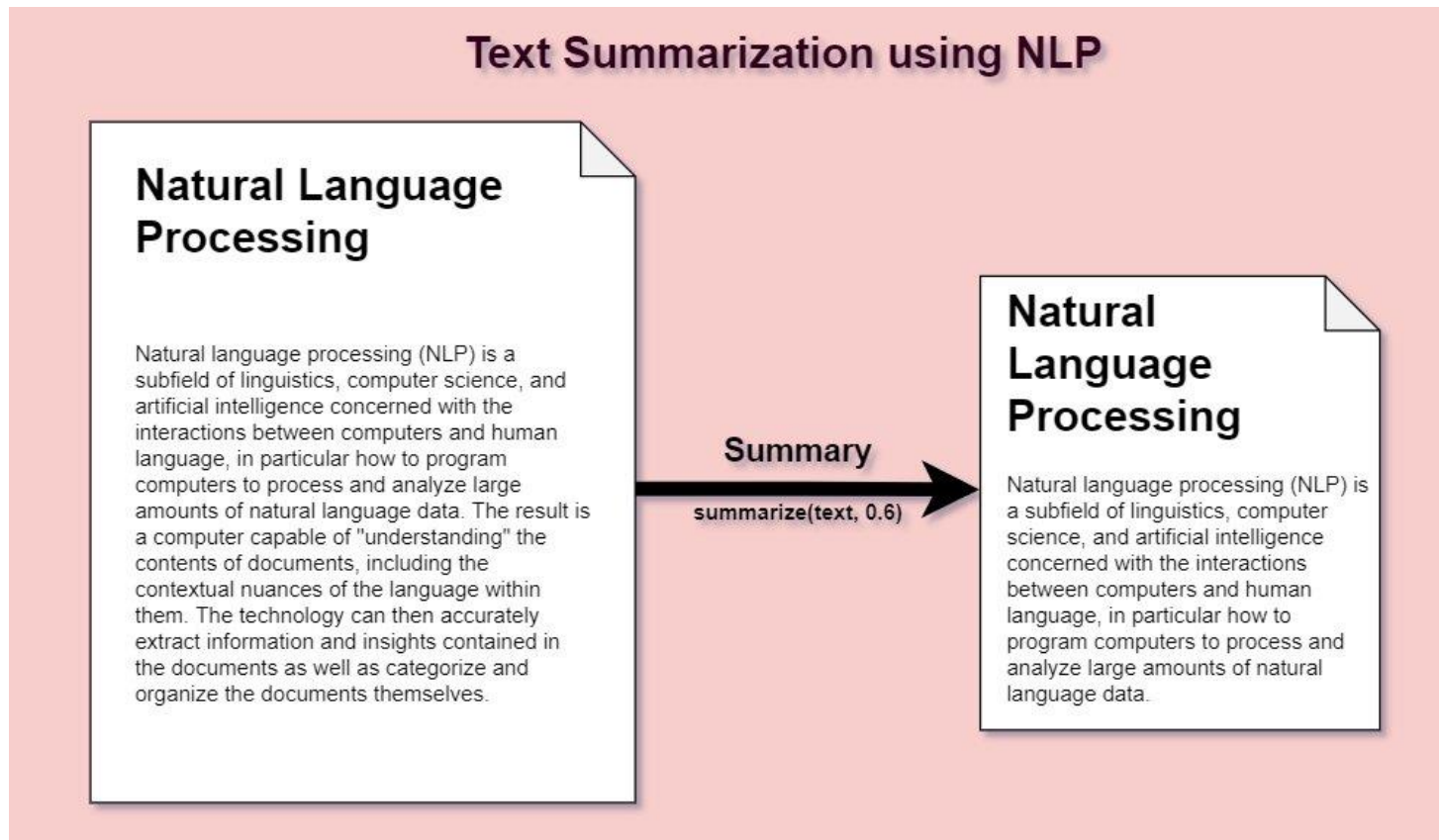
What is
Automatic Text
Summarization

Methods

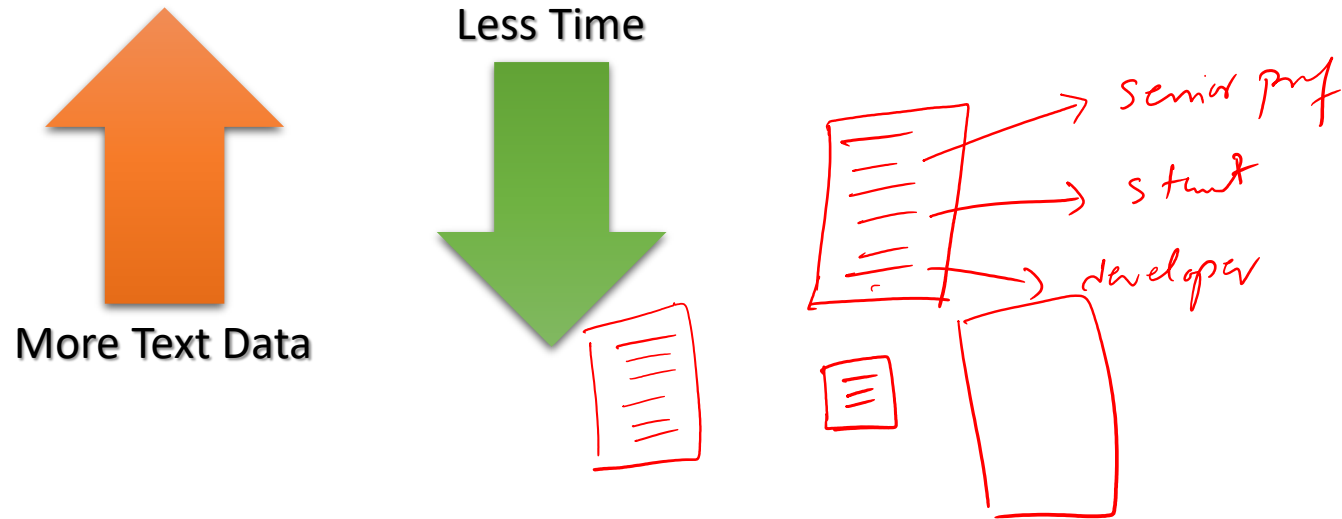
Evaluation

What is Automatic Text Summarization?

Automatic summarization is the process of shortening a set of data computationally, to create a subset (a summary) that represents the most important or relevant information within the original content. (Wikipedia)



Why Summary?



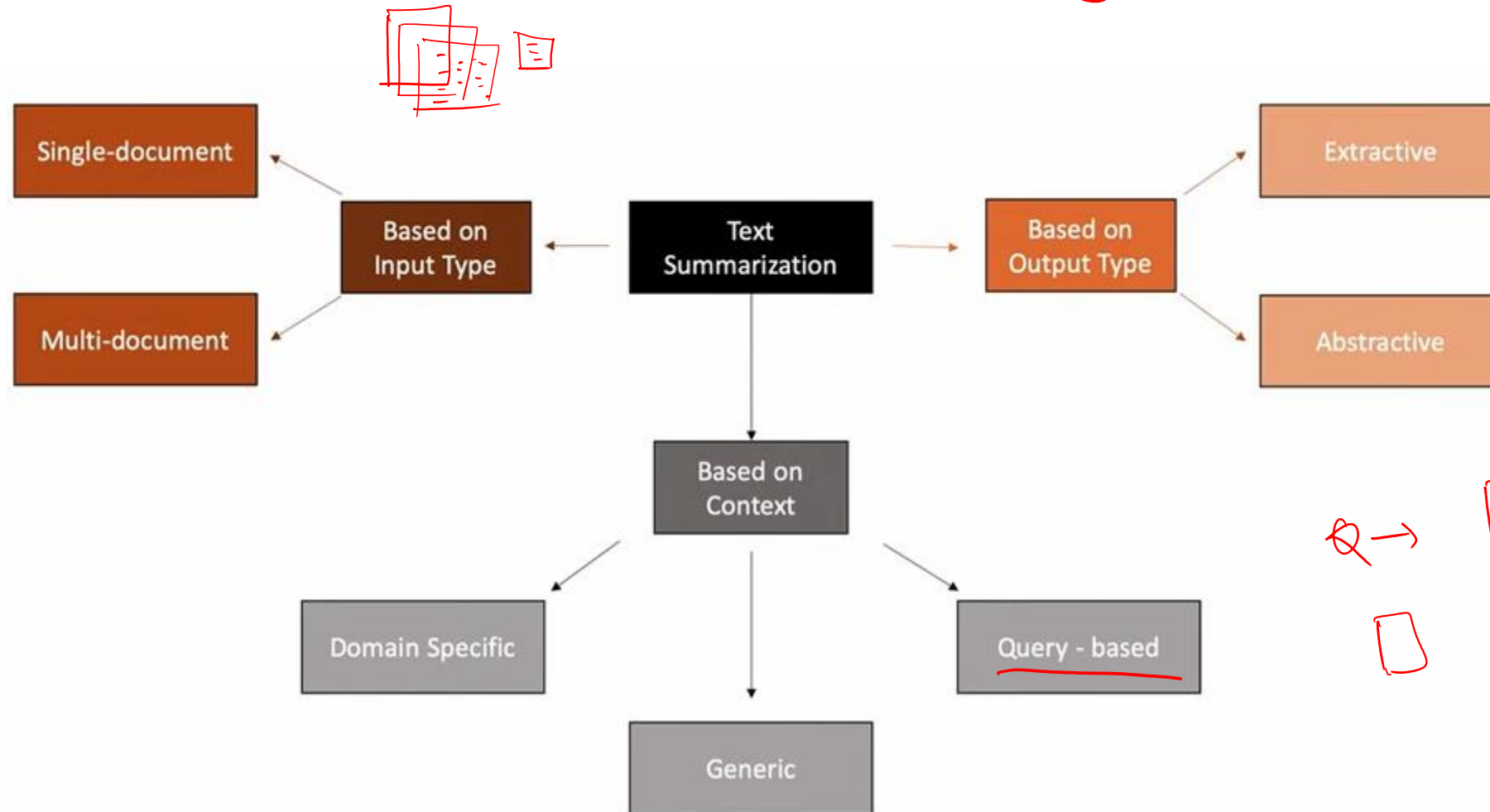
Why Difficult?

- Selecting the most relevant information from a source document
- Expressing that key information in the final summary

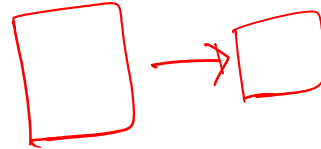
Objective

- How can you identify the most important thing in the shortest amount of time
 - Optimize information coverage
 - Optimize readability

Type of text summarization



Extractive vs Abstractive Summarization



Extractive

Kent Sprouse was put to death by lethal injection Thursday ... that he killed a police officer and another man outside a Dallas-area convenience store... decided in 2004 that ...

Abstractive

In 2004, Kent Sprouse received a death sentence for murders a decade ago.

Kent Sprouse was put to death by lethal injection Thursday night in Texas. Sprouse acknowledged almost immediately after he was arrested more than a decade ago that he killed a police officer and another man outside a Dallas-area convenience store. A jury decided in 2004 that he should be put to death, and unsuccessful appeals since then have focused on whether Sprouse was mentally ill at the time of the slayings in 2002 and should be spared the death penalty.

Extractive Summarization

Game of Thrones is an American fantasy drama television series created by David Benioff and D. B. Weiss for HBO. It is an adaptation of A Song of Ice and Fire, George R. R. Martin's series of fantasy novels, the first of which is A Game of Thrones. The show was both produced and filmed in Belfast and elsewhere in the United Kingdom. Filming locations also included Canada, Croatia, Iceland, Malta, Morocco, and Spain. The series premiered on HBO in the United States on April 17, 2011, and concluded on May 19, 2019, with 73 episodes broadcast over eight seasons.

Set on the fictional continents of Westeros and Essos, Game of Thrones has several plots and a large ensemble cast and follows several story arcs. One arc is about the Iron Throne of the Seven Kingdoms and follows a web of alliances and conflicts among the noble dynasties either vying to claim the throne or fighting for independence from it. Another focuses on the last descendant of the realm's deposed ruling dynasty, who has been exiled and is plotting a return to the throne, while another story arc follows the Night's Watch, a brotherhood defending the realm against the fierce peoples and legendary creatures of the North.

Advantage: Grammatical correctness


Extractive Summarization-Lack of Balance

Game of Thrones is an American fantasy drama television series created by David Benioff and D. B. Weiss for HBO. It is an adaptation of A Song of Ice and Fire, George R. R. Martin's series of fantasy novels, the first of which is A Game of Thrones. The show was both produced and filmed in Belfast and elsewhere in the United Kingdom. Filming locations also included Canada, Croatia, Iceland, Malta, Morocco, and Spain. The series premiered on HBO in the United States on April 17, 2011, and concluded on May 19, 2019, with 73 episodes broadcast over eight seasons.

(Set on the fictional continents of Westeros and Essos, Game of Thrones has several plots and a large ensemble cast and follows several story arcs.) One arc is about the Iron Throne of the Seven Kingdoms and follows a web of alliances and conflicts among the noble dynasties either vying to claim the throne or fighting for independence from it. Another focuses on the last descendant of the realm's deposed ruling dynasty, who has been exiled and is plotting a return to the throne, while another story arc follows the Night's Watch, a brotherhood defending the realm against the fierce peoples and legendary creatures of the North.

Extractive Summarization-Lack of Balance

Game of Thrones is an American fantasy drama television series created by David Benioff and D. B. Weiss for HBO. It is an adaptation of A Song of Ice and Fire, George R. R. Martin's series of fantasy novels, the first of which is A Game of Thrones. The show was both produced and filmed in Belfast and elsewhere in the United Kingdom. Filming locations also included Canada, Croatia, Iceland, Malta, Morocco, and Spain. The series premiered on HBO in the United States on April 17, 2011, and concluded on May 19, 2019, with 73 episodes broadcast over eight seasons.



Set on the fictional continents of Westeros and Essos, Game of Thrones has several plots and a large ensemble cast and follows several story arcs. One arc is about the Iron Throne of the Seven Kingdoms and follows a web of alliances and conflicts among the noble dynasties either vying to claim the throne or fighting for independence from it. Another focuses on the last descendant of the realm's deposed ruling dynasty, who has been exiled and is plotting a return to the throne, while another story arc follows the Night's Watch, a brotherhood defending the realm against the fierce peoples and legendary creatures of the North.

Extractive Summarization-Lack of Cohesion

Game of Thrones is an American fantasy drama television series created by David Benioff and D. B. Weiss for HBO. It is an adaptation of A Song of Ice and Fire, George R. R. Martin's series of fantasy novels, the first of which is A Game of Thrones. The show was both produced and filmed in Belfast and elsewhere in the United Kingdom. Filming locations also included Canada, Croatia, Iceland, Malta, Morocco, and Spain. The series premiered on HBO in the United States on April 17, 2011, and concluded on May 19, 2019, with 73 episodes broadcast over eight seasons.

Set on the fictional continents of Westeros and Essos, Game of Thrones has several plots and a large ensemble cast and follows several story arcs. One arc is about the Iron Throne of the Seven Kingdoms and follows a web of alliances and conflicts among the noble dynasties either vying to claim the throne or fighting for independence from it. Another focuses on the last descendant of the realm's deposed ruling dynasty, who has been exiled and is plotting a return to the throne, while another story arc follows the Night's Watch, a brotherhood defending the realm against the fierce peoples and legendary creatures of the North.

Abstractive Summarization

~~Game of Thrones~~ is an American fantasy drama television series created by David Benioff and D. B. Weiss for HBO. It is an adaptation of A Song of Ice and Fire, George R. R. Martin's series of fantasy novels, the first of which is A Game of Thrones. The show was both produced and filmed in Belfast and elsewhere in the United Kingdom. Filming locations also included Canada, Croatia, Iceland, Malta, Morocco, and Spain. The series premiered on HBO in the United States on April 17, 2011, and concluded on May 19, 2019, with 73 episodes broadcast over eight seasons.

Set on the fictional continents of Westeros and Essos, Game of Thrones has several plots and a large ensemble cast and follows several story arcs. One arc is about the Iron Throne of the Seven Kingdoms and follows a web of alliances and conflicts among the noble dynasties either vying to claim the throne or fighting for independence from it. Another focuses on the last descendant of the realm's deposed ruling dynasty, who has been exiled and is plotting a return to the throne, while another story arc follows the Night's Watch, a brotherhood defending the realm against the fierce peoples and legendary creatures of the North.

Game of Thrones is a TV show based on book series A Song of Ice and Fire, written by G. R. R. Martin. All eight seasons were filmed in many beautiful countries across three different continents. Game of Thrones has a very complex story with several plots and story arcs – from conflicts between Westeros nobility to claim the Iron Throne and rule over Seven Kingdoms to fight between brotherhood called Night's watch and enemies from the North.

What is
Automatic Text
Summarization

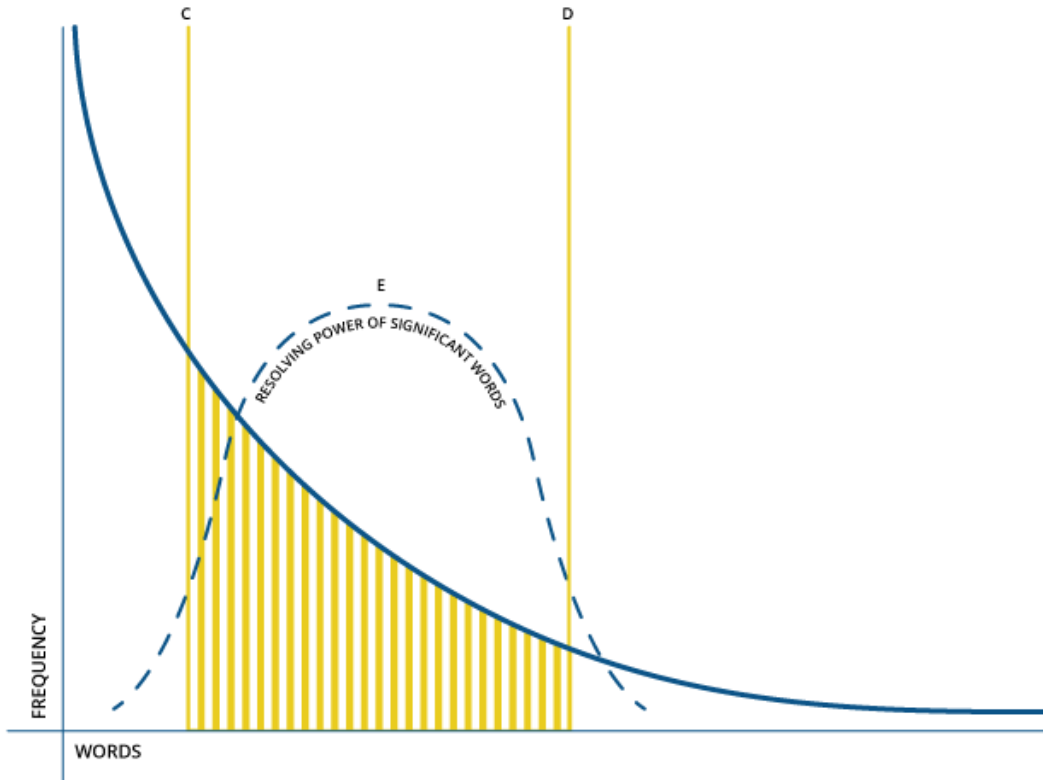
Methods

Evaluation

Positional Method (Baxendale, 1958)

- Technical literature
- Analysis on 200 paragraphs
- First and last sentence of a paragraph are topic sentences (85% and 7%)

Luhl's Method, 1958



- Extract salient sentences from the text using features such as word and phrase frequency.
- Weight the sentences of a document as a function of high frequency words, ignoring very high frequency common words.

Tf-idf

Edmundson's method

- Introduced by H. P. Edmundson in 1968
 - «*New Methods in Automatic Extracting* »
- Position (P)
- Word frequency (F)
- Cue words (C)
 - **Bonus words** – pointing to the important sentence
 - **Stigma words** – negative effect on the sentence importance
 - **Null words** – neutral or irrelevant to the importance of the sentence
- Document structure (S)
- Linear combination of these 4 features: $\alpha_1 P + \alpha_2 F + \alpha_3 C + \alpha_4 S$

Other methods

- Topic representations
- Word Probability
- TF-IDF
- Topic word Approaches
- Machine-Learning Methods

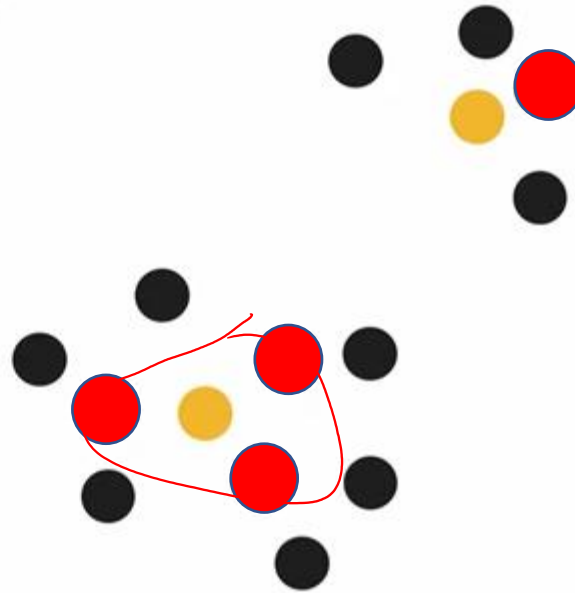
Mead

- Introduced by Radev et al. in 2000
- Centroid-based method
- Single and multi document



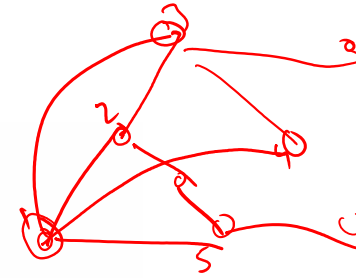
Mead

- Introduced by Radev et al. in 2000
- Centroid-based method
- Single and multi document



Graph Based Approaches

LexRank



- Introduced by Erkan and Radev in 2004

- «LexRank: Graph-based Lexical Centrality as Saliency in Text Summarization»

- Graph based method

- **Lexical centrality**

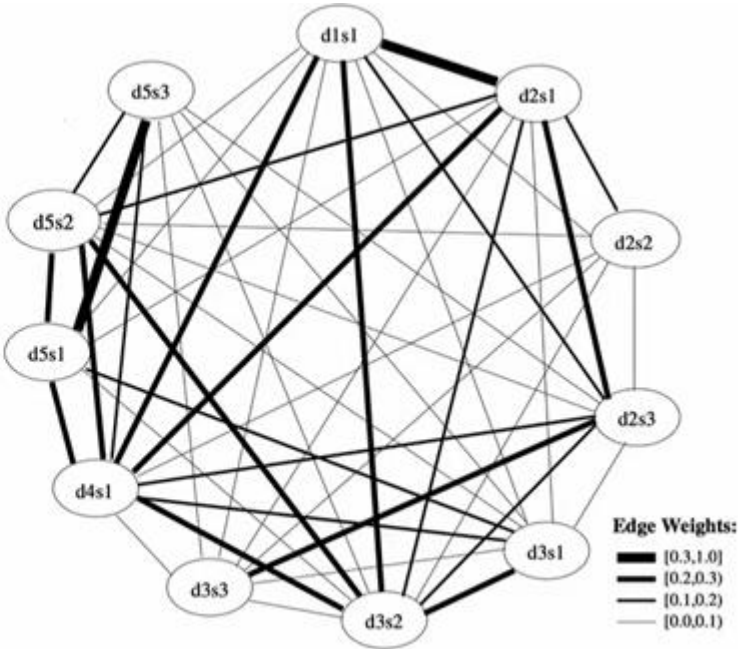
- Idea:

- Create a similarity matrix
 - Representing sentences as nodes in the graph
 - Connecting nodes based on similarity matrix

1	d1s1	¹ Iraqi Vice President Taha Yassin Ramadan announced today, Sunday, that Iraq refuses to back down from its decision to stop cooperating with disarmament inspectors before its demands are met.
2	d2s1	Iraqi Vice president Taha Yassin Ramadan announced today, Thursday, that Iraq rejects cooperating with the United Nations except on the issue of lifting the blockade imposed upon it since the year 1990.
3	d2s2	Ramadan told reporters in Baghdad that "Iraq cannot deal positively with whoever represents the Security Council unless there was a clear stance on the issue of lifting the blockade off of it.
4	d2s3	Baghdad had decided late last October to completely cease cooperating with the inspectors of the United Nations Special Commission (UNSCOM), in charge of disarming Iraq's weapons, and whose work became very limited since the fifth of August, and announced it will not resume its cooperation with the Commission even if it were subjected to a military operation.
5	d3s1	The Russian Foreign Minister, Igor Ivanov, warned today, Wednesday against using force against Iraq, which will destroy, according to him, seven years of difficult diplomatic work and will complicate the regional situation in the area.
6	d3s2	Ivanov contended that carrying out air strikes against Iraq, who refuses to cooperate with the United Nations inspectors, "will end the tremendous work achieved by the international group during the past seven years and will complicate the situation in the region."
7	d3s3	Nevertheless, Ivanov stressed that Baghdad must resume working with the Special Commission in charge of disarming the Iraqi weapons of mass destruction (UNSCOM).
8	d4s1	The Special Representative of the United Nations Secretary-General in Baghdad, Prakash Shah, announced today, Wednesday, after meeting with the Iraqi Deputy Prime Minister Tariq Aziz, that Iraq refuses to back down from its decision to cut off cooperation with the disarmament inspectors.
9	d5s1	British Prime Minister Tony Blair said today, Sunday, that the crisis between the international community and Iraq "did not end" and that Britain is still "ready, prepared, and able to strike Iraq."
10	d5s2	In a gathering with the press held at the Prime Minister's office, Blair contended that the crisis with Iraq "will not end until Iraq has absolutely and unconditionally respected its commitments" towards the United Nations
11	d5s3	A spokesman for Tony Blair had indicated that the British Prime Minister gave permission to British Air Force Tornado planes stationed in Kuwait to join the aerial bombardment against Iraq.

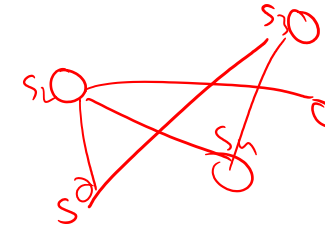
Graph Based Approaches

	1	2	3	4	5	6	7	8	9	10	11
1	1.00	0.45	0.02	0.17	0.03	0.22	0.03	0.28	0.06	0.06	0.00
2	0.45	1.00	0.16	0.27	0.03	0.19	0.03	0.21	0.03	0.15	0.00
3	0.02	0.16	1.00	0.03	0.00	0.01	0.03	0.04	0.00	0.01	0.00
4	0.17	0.27	0.03	1.00	0.01	0.16	0.28	0.17	0.00	0.09	0.01
5	0.03	0.03	0.00	0.01	1.00	0.29	0.05	0.15	0.20	0.04	0.18
6	0.22	0.19	0.01	0.16	0.29	1.00	0.05	0.29	0.04	0.20	0.03
7	0.03	0.03	0.03	0.28	0.05	0.05	1.00	0.06	0.00	0.00	0.01
8	0.28	0.21	0.04	0.17	0.15	0.29	0.06	1.00	0.25	0.20	0.17
9	0.06	0.03	0.00	0.00	0.20	0.04	0.00	0.25	1.00	0.26	0.38
10	0.06	0.15	0.01	0.09	0.04	0.20	0.00	0.20	0.26	1.00	0.12
11	0.00	0.00	0.00	0.01	0.18	0.03	0.01	0.17	0.38	0.12	1.00



Graph Based Approaches

- TextRank ([Pytextrank](#))
 - Motivated by google pagerank
 - Document as a graph

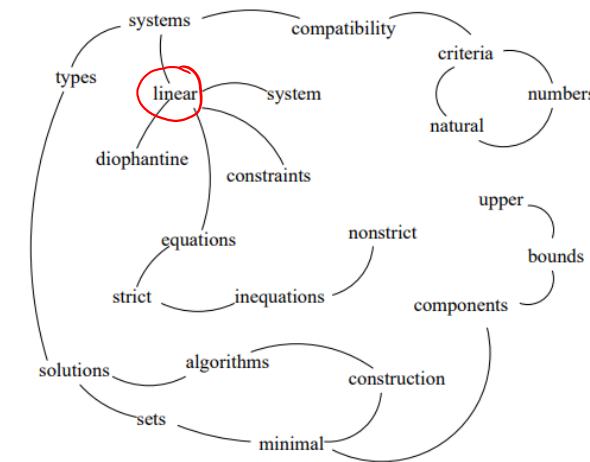
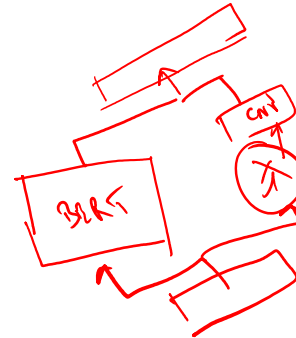
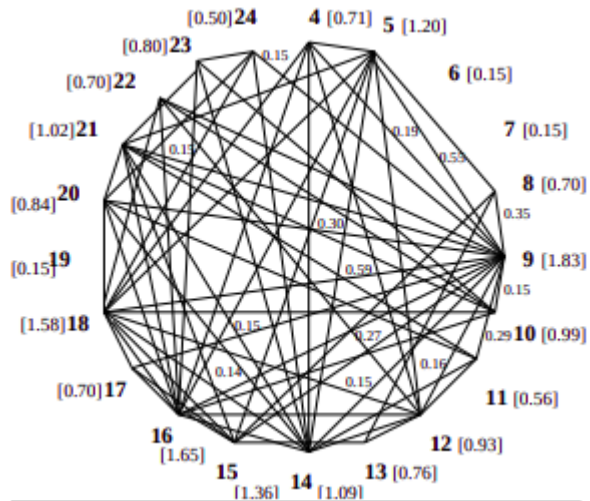


$$S(V_i) = (1 - d) + d * \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j)$$

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j)$$

Compatibility of systems of linear constraints over the set of natural numbers. Criteria of compatibility of a system of linear Diophantine equations, strict inequations, and nonstrict inequations are considered. Upper bounds for components of a minimal set of solutions and algorithms of construction of minimal generating sets of solutions for all types of systems are given. These criteria and the corresponding algorithms for constructing a minimal supporting set of solutions can be used in solving all the considered types systems and systems of mixed types.

- Extracting Keywords/keyphrases
- Sentence extraction task



Keywords assigned by TextRank:

linear constraints; linear diophantine equations; natural numbers; nonstrict inequations; strict inequations; upper bounds

Keywords assigned by human annotators:

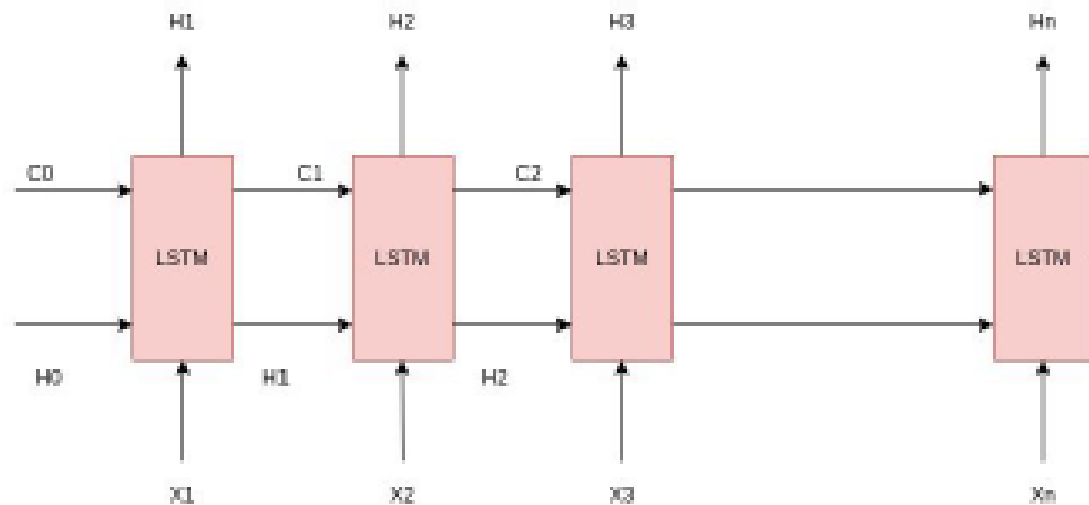
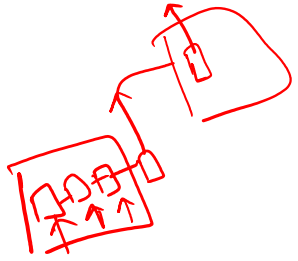
linear constraints; linear diophantine equations; minimal generating sets; non-strict inequations; set of natural numbers; strict inequations; upper bounds

Abstractive Summarization

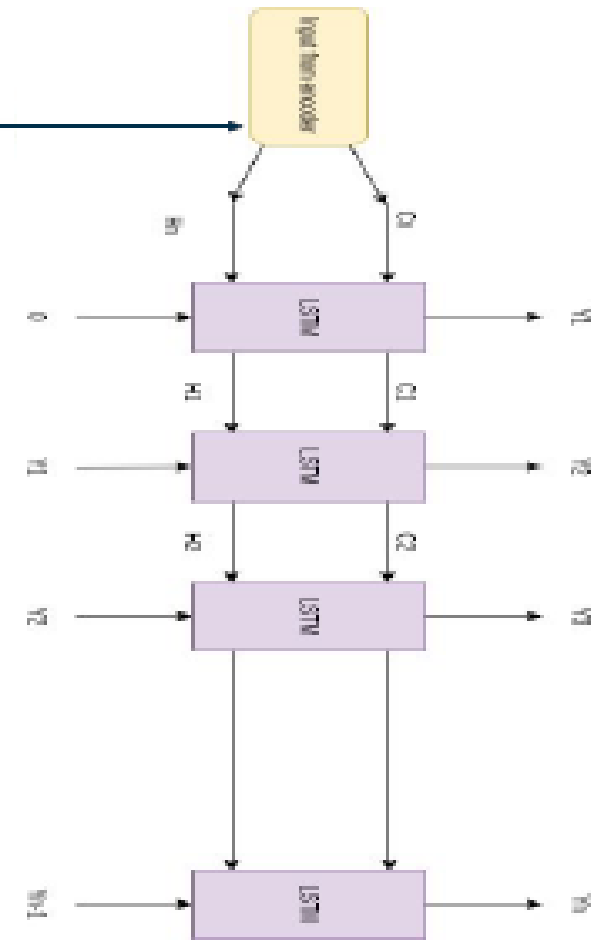
- Does not simply select sentences from the originally given text passage to
- create the summary
- They produce a *paraphrasing of the main contents of the given text* - using a vocabulary set possibly different from the original document
- Creates a semantic representation of the long document – generates another shorter representation that is semantically equivalent to the earlier one
- Involves - Natural Language Generation

$w_1, w_2, w_3, w_4 \rightarrow w_x, w_y, w_z, \dots$

Encoder - Decoder architecture using RNNs



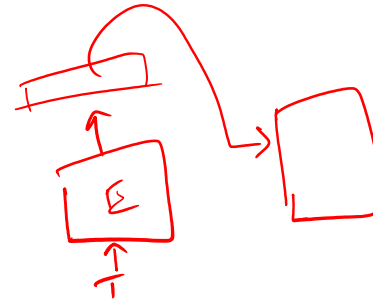
- The encoder-decoder model is trained on a target set or vocabulary of words
- The LSTMs hidden activation is sent through a softmax layer which generates the probabilities for each word in the vocabulary to be predicted as the next word.
- The word with the maximum probability is chosen as the output at that time step.
- To know which word exactly suits the semantics - the model trains on a dataset and solves the problem as a supervised classification problem.



Issue with LSTM based Encoder-Decoder

- If the original document is very large – the focus of the model is more on the words towards the end, not what is in the beginning
- Humans on the other hand usually summarize from 1–5 lines
 - Moves forward slowly
 - Rarely goes up to consider the 100th line
- This aspect could not be implemented by the normal encoder-decoder model

Transformer based Models



- BART¹

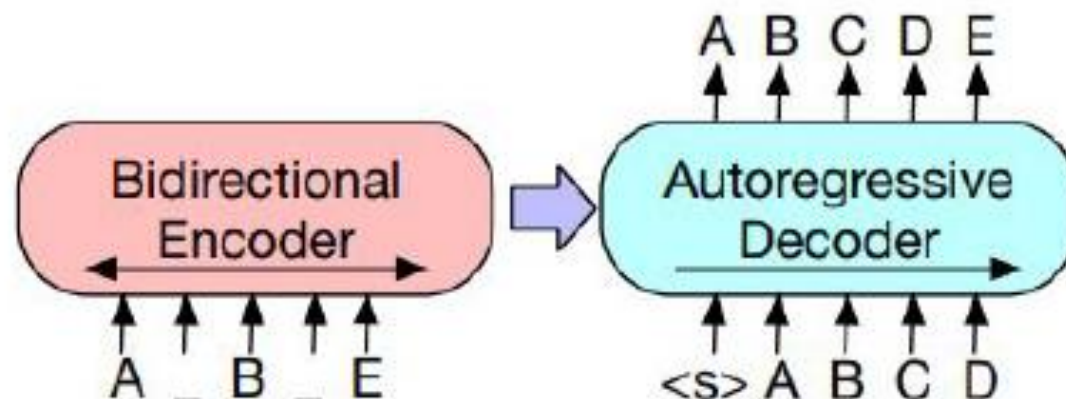
- Sequence-to-Sequence - Pre-training for Natural Language Generation, Translation, and Comprehension – 2019

- PEGASUS

- PEGASUS: pre-training with extracted Gap-sentences for Abstractive summarization – 2020

BART Model

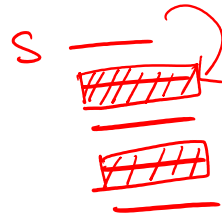
- Pre-trains a model cc Transformers



- BART is trained by corrupting documents and then optimizing a reconstruction loss—the cross-entropy between the decoder's output and the original document
- BART allows users to apply any type of document corruption

PEGASUS

- Important sentences are removed/masked from an input document and are generated together as one output sequence from the remaining sentences, similar to an extractive summary.



PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization :
Proceedings of the 37 th International Conference on Machine Learning, Online, PMLR 119,
2020)

Pretraining objectives were meant for summarization tasks

- For abstractive text summarization and evaluate on 12 downstream datasets spanning news
- Masking whole sentences from a document and generating these **gap-sentences** from the rest of the document works well as a pre-training objective for downstream summarization tasks
- PEGASUS masks multiple whole sentences rather than smaller continuous text spans
- In particular, choosing putatively important sentences outperforms lead or randomly selected ones

Evaluation

What makes a good summary?

○ Goals:

- Optimize topic coverage
- Optimize readability

○ Evaluation criteria:

- Salience
- Length
- Structure and coherence
- Balance
- Grammar
- Non-redundancy

Information content

$$C = \text{Compression Ratio} = \frac{|S|}{|D|}$$

$$R = \text{Retention Ratio} = \frac{i(S)}{i(D)}$$



- we want to optimize this information content basically what we want is to optimize the compression and retention ratio.
- where **compression ratio** is just telling us okay what is the percentage of the original document that ended up in the summary and
- **retention ratio** is telling us okay what is the percentage of information of the original document that was preserved in the summary

Personalized Summarization..

Precision and Recall

100

100

	Ideal	Summarizer 1	Summarizer 2
S1	+	+	+
S2	+	+	-
S3	-	-	-
S4	-	-	+
S5	-	-	-
S6	-	-	-
S7	-	-	-
S8	-	-	-
S9	-	-	-
S10	-	-	-

100 [] 0 0 0 0

	Manually selected	Manually not selected
Automatic selected	True positives	False positives
Automatic not selected	False negatives	True negatives

$$\text{Precision} = \frac{\text{True positives}}{\text{False positives} + \text{True positives}}$$

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}}$$

Utility Based on ranking of the sentences

	Ideal	Summarizer 1	Summarizer 2
10 → S1	+	+	+
5 → S2	+	+	-
S3	-	-	-
S4	-	-	+
S5	-	-	-
S6	-	-	-
S7	-	-	-
S8	-	-	-
S9	-	-	-
S10	-	-	-

S3

-

+

+

	Ideal	Summarizer 1	Summarizer 2
S1	10 (+)	+	+
S2	9 (+)	+	-
S3	1 (-)	-	-
S4	8 (-)	-	+
S5	2 (-)	-	-

BLEU Score (Mainly for MTs)

- ~~$BLEU = BP * \exp(\sum pn)$~~
- BP (Brevity Penalty)
- $BP = \min(1, \frac{\text{reference_length}}{\text{translated_length}})$
- pn is the precision of n-grams
- Number of n-grams that appear in both the machine-generated translation and the reference translations divided by the total number of n-grams in the machine-generated translation.
- Ranges from 0 to 1
- higher values → better translation quality.
- A perfect translation would have a BLEU score of 1, while a completely incorrect translation would have a BLEU score of 0.

7
10

Recall-Oriented Understudy for Gisting Evaluation

- ROUGE = \sum (Recall of n-grams)
- Recall of n-grams is the number of n-grams that appear in both the machine-generated summary and the reference summaries divided by the total number of n-grams in the reference summaries.
- Ranges from 0 to 1
- higher values \rightarrow better summary quality.
- Like BLEU score, a perfect summary would have a ROUGE score of 1, while a completely incorrect summary would have a ROUGE score of 0.
- ROUGE scores are branched into ROUGE-N, ROUGE-L, and ROUGE-S.

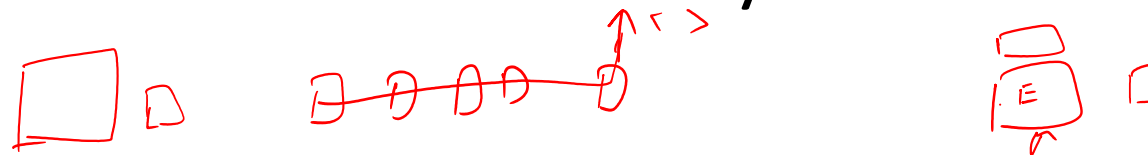
$$R = \sum \text{Recall of } n\text{-grams}$$

ROUGE-1

= no of overlapping unigrams

R-2 = " " " " Bigram.

Recall-Oriented Understudy for Gisting Evaluation



- ROUGE-N: measures the overlap of n-grams (contiguous sequences of n words) between the candidate text and the reference text.
- It computes the precision, recall, and F1-score based on the n-gram overlap.
- For example,
 - ROUGE-1 (unigram) measures the overlap of single words,
 - ROUGE-2 (bigram) measures the overlap of two-word sequences, and so on.
 - ROUGE-N is often used to evaluate the grammatical correctness and fluency of generated text.
- ROUGE-L: measures the longest common subsequence (LCS) between the candidate text and the reference text.
- It computes the precision, recall, and F1-score based on the length of the LCS.
- ROUGE-L is often used to evaluate the semantic similarity and content coverage of generated text, as it considers the common subsequence regardless of word order.

→ S: Summarization is cool
 → R: Summarization is beneficial & cool.

{S is}, (is cool)
 (S is), (is bene), (benef) (of cool)

S is cool but different
 R1 → 'S', 'is', 'cool'
 → S is B & cool.

$$R = \frac{3}{5}$$

$$P = \frac{3}{3} = 1$$

$$\frac{3}{5} = .6$$

Evaluation

- ROUGE: ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation.

- ROUGE-n

- ROUGE-L

$$ROUGE-2 = \frac{\sum_{s \in \{\text{RefSummaries}\}} \sum_{\text{bigrams } i \in S} \min(\text{count}(i, X), \text{count}(i, S))}{\sum_{s \in \{\text{RefSummaries}\}} \sum_{\text{bigrams } i \in S} \text{count}(i, S)}$$

X: [A B C D E F G]
Y: [A B C D H I K]
Y: [A H B K C I D]

Limitations

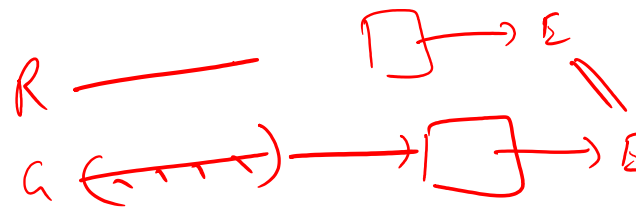
- BLEU score, ROUGE score also has limitations.
- It may not fully capture the semantic meaning or coherence of the summary, and it relies solely on the n-gram overlap, which may not always be an accurate measure of summary quality.
- Difficult to use for Abstractive Summarization

METEOR (Metric for Evaluation of Translation with Explicit ORdering)

- Exact word matching: Exact matches between reference and generation
- ~~Stemmed matching~~: Exact matches after Porter stemming
- ~~Synonym matching~~: Matches via WordNet synonyms
- Paraphrase matching: Matches on phrases that are structured differently

R —
G —
—
—
—

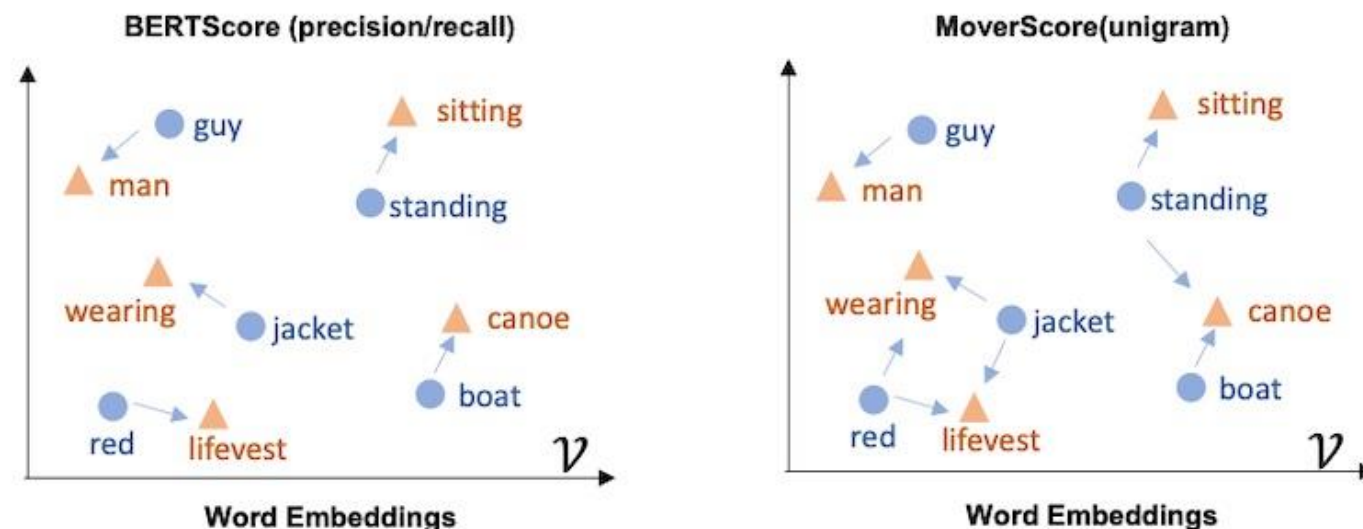
BERTScore



- It is an embedding-based metric that extends on the spirit of METEOR.
 - uses cosine similarity to compare each token or n-gram in the generated summary with the reference.
- There are three components to BERTScore:
 - Precision: Average cosine similarity between each token in the generated output and its nearest match in the reference.
 - Recall: Average cosine similarity between each token in the reference and its closest match in the generated output.
 - F1: Harmonic mean of precision and recall

MoverScore

- Also uses contextualized embeddings to compute the distance between tokens in the generated output and reference.
- But unlike BERTScore, which is based on one-to-one matching (or “high alignment”) of tokens, MoverScore allows for many-to-one matching (or “soft alignment”).



- System x: A guy with a red jacket is standing on a boat
- ▲ Ref y: A man wearing a lifevest is sitting in a canoe

Figure 1: An illustration of MoverScore and BERTScore.

Other Metrics

1. **Relevance:** Evaluates if the summary includes only important information and excludes redundancies.
2. **Coherence:** Assesses the logical flow and organization of the summary.
3. **Consistency:** Checks if the summary aligns with the facts in the source document.
4. **Fluency:** Rates the grammar and readability of the summary.

Hallucination

Summary of Summarizers

Tools



○ **sumy library**

- Luhn
- Edmundson
- LSA
- LexRank
- TextRank
- SumBasic
- KL-Sum
- Reduction