

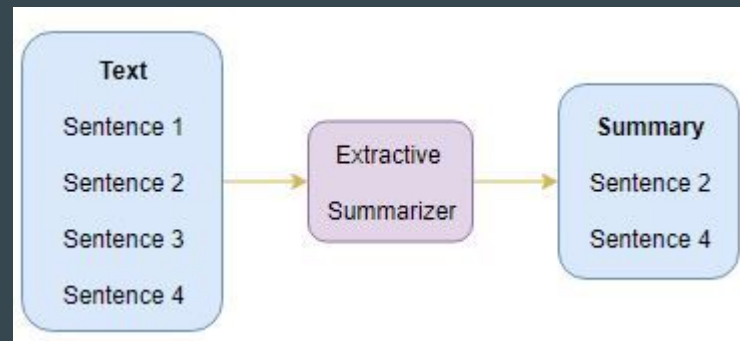
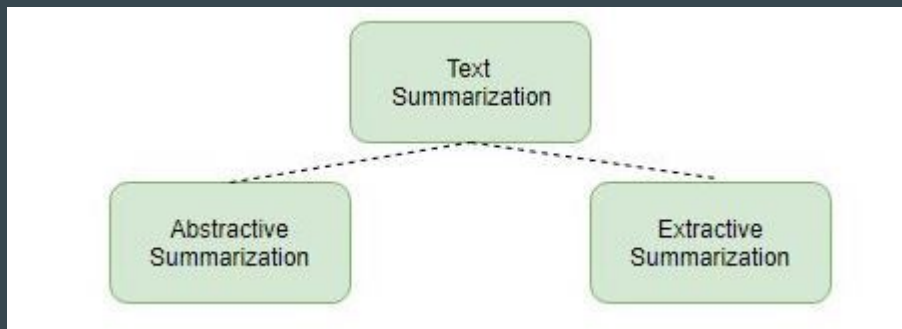
Graph-based text summarization

...

Julia Rodina

About

- Task: extractive summarization of news articles
- Data: Russian 2019 news crawled from Meduza
- Method: TextRank variations

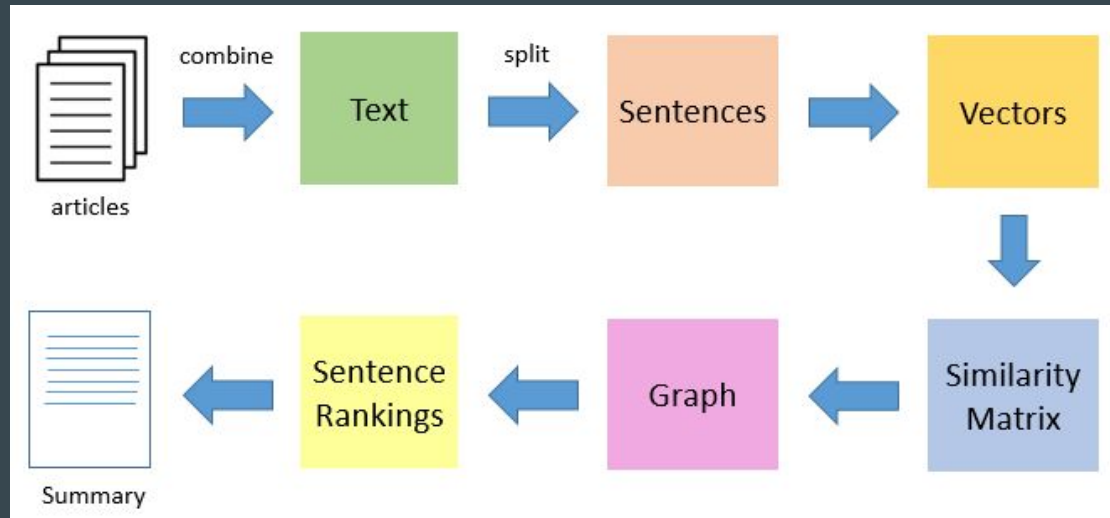


TextRank

- In place of web pages, we use sentences
- Similarity between any two sentences is used as an equivalent to the web page transition probability
- The similarity scores are stored in a square matrix, similar to the matrix used for PageRank

webpage	links
w1	[w4, w2]
w2	[w3, w1]
w3	[]
w4	[w1]

M =		w1	w2	w3	w4
	w1	0	0.5	0	0.5
	w2	0.5	0	0.5	0
	w3	0.25	0.25	0.25	0.25
	w4	1	0	0	0



Method 1 (basic)

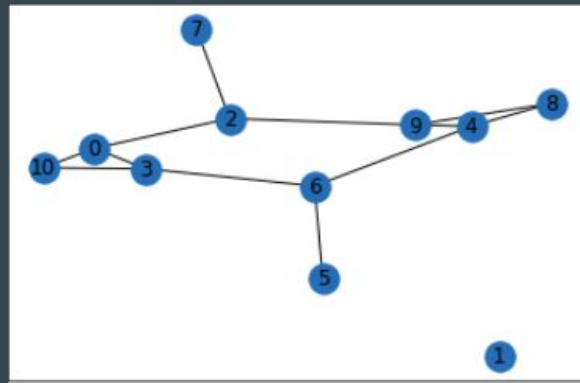
Input article → split into sentences → remove stop words → build a similarity matrix → generate rank based on matrix → pick top N sentences for summary.

Input text:

11 тысяч ученых и экологов выпустили обращение и доклад под заголовком «Предупреждение ученых мира о климатической чрезвычайной ситуации» Оно опубликовано в британском научном журнале BioScience в разделе «Мнения» Земля уже находится в «чрезвычайной климатической ситуации», говорится в обращении, приуроченном к 40-летию первой Всемирной климатической конференции в Женеве в 1979 году «Климатический кризис наступил и усугубляется быстрее, чем ожидало большинство ученых Он оказался более серьезным, чем ожидалось, и угрожает природным экосистемам и судьбе человечества», – заявили авторы обращения Они призвали установить контроль за «ключевыми индикаторами» причин и последствий изменения климата, включив в их число рост населения, потребление мяса, исчезновение зеленого покрова, потребление энергии и еще ряд показателей Авторы также заявили, что климатический кризис тесно связан с чрезмерным потреблением, включая, к примеру, стремительный рост пассажирских авиаперевозок Результатом стало повышение температуры суши и океана, повышение уровня моря и участвовавшие экстремальные погодные явления, говорится в обращении Ведущим соавтором обращения стал профессор Уильям Риппл из Орегонского государственного университета В 2017 году он же был инициатором обращения об опасности загрязнения планеты и массовом исчезновении диких видов на Земле Его подписали более 15 тысяч ученых Обращение было опубликовано в качестве продолжения письма ученых от 1992 года об исчезновении озонового слоя, лесов, чистой воды и почвы на планете.

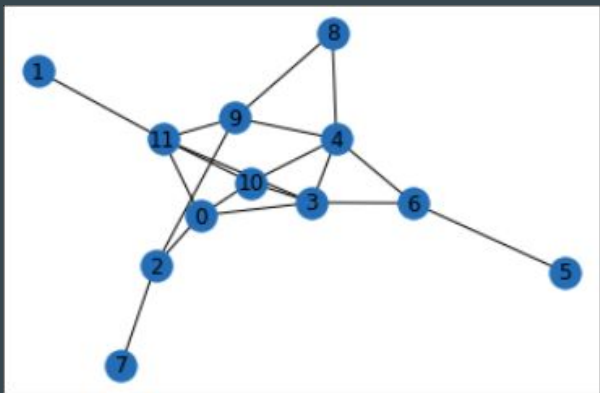
Summarized Text:

11 тысяч ученых и экологов выпустили обращение и доклад под заголовком «Предупреждение ученых мира о климатической чрезвычайной ситуации». Его подписали более 15 тысяч ученых. «Климатический кризис наступил и усугубляется быстрее, чем ожидало большинство ученых



Method 2

With sklearn CountVectorizer and
TfidfTransformer.



Summary:

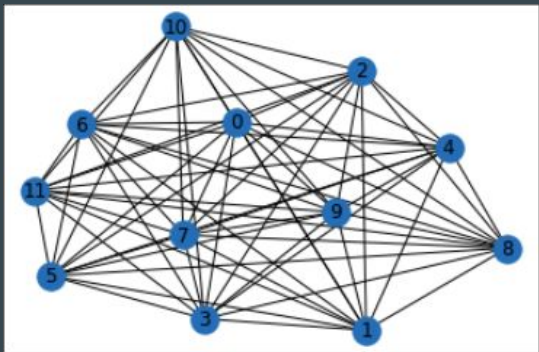
11 тысяч ученых и экологов выпустили обращение и доклад под заголовком «Предупреждение ученых мира о климатической чрезвычайной ситуации».

Он оказался более серьезным, чем ожидалось, и угрожает природным экосистемам и судьбе человечества», — заявили авторы обращения.

Обращение было опубликовано в качестве продолжения письма ученых от 1992 года об исчезновении озонового слоя, лесов, чистой воды и почвы на планете.

Method 3

With pretrained fasttext embeddings



Summary:

11 тысяч ученых и экологов выпустили обращение и доклад под заголовком «Предупреждение ученых мира о климатической чрезвычайной ситуации». Обращение было опубликовано в качестве продолжения письма ученых от 1992 года об исчезновении озонового слоя, лесов, чистой воды и почвы на планете. Они призвали установить контроль за «ключевыми индикаторами» причин и последствий изменения климата, включив в их число рост населения, потребление мяса, исчезновение зеленого покрова, потребление энергии и еще ряд показателей.

Method 4

SummaNLP library



Summary with ratio=0.3:

Он оказался более серьезным, чем ожидалось, и угрожает природным экосистемам и судьбе человечества», — заявили авторы обращения. 'В 2017 году он же был инициатором обращения об опасности загрязнения планеты и массовом исчезновении диких видов на Земле. 'Обращение было опубликовано в качестве продолжения письма ученых от 1992 года об исчезновении озонового слоя, лесов, чистой воды и почвы на планете.

Evaluation

1. Text average vectors cosine similarity
2. Keywords extraction

1. Elmo text embeddings, results are different. For given example:

summary1: [[0.8172235]]; summary2: [[0.90667229]]

summary3: [[0.92517671]]; summary4: [[0.89708047]]

2.

Intersection of text and summary1 keywords: {'ученых'}

Intersection of text and summary2 keywords: {'ученых', 'обращение'}

Intersection of text and summary3 keywords: {'исчезновение', 'обращение', 'ученых', 'исчезновении', 'потребление'}

Intersection of text and summary4 keywords: {'года', 'обращение', 'планеты', 'исчезновении', 'планете'}



Any questions?