

Phramer → TL; DR

Быстрый улов сути текста

Каждую секунду появляются:

6K твитов

3.5M и-мейлов

11 новых пользователей интернета

Источник: <https://www.visualcapitalist.com/how-much-data-is-generated-each-day/>



Верхний левый угол — Николай Шаталов; верхний правый угол — Вадим Филиппович, куратор команды; левый нижний угол — Саша Илларионов; нижний правый угол — Паша Факанов.

Команда

Павел Факанов

ФКН ПМИ, 3 курс

Саша Илларионов

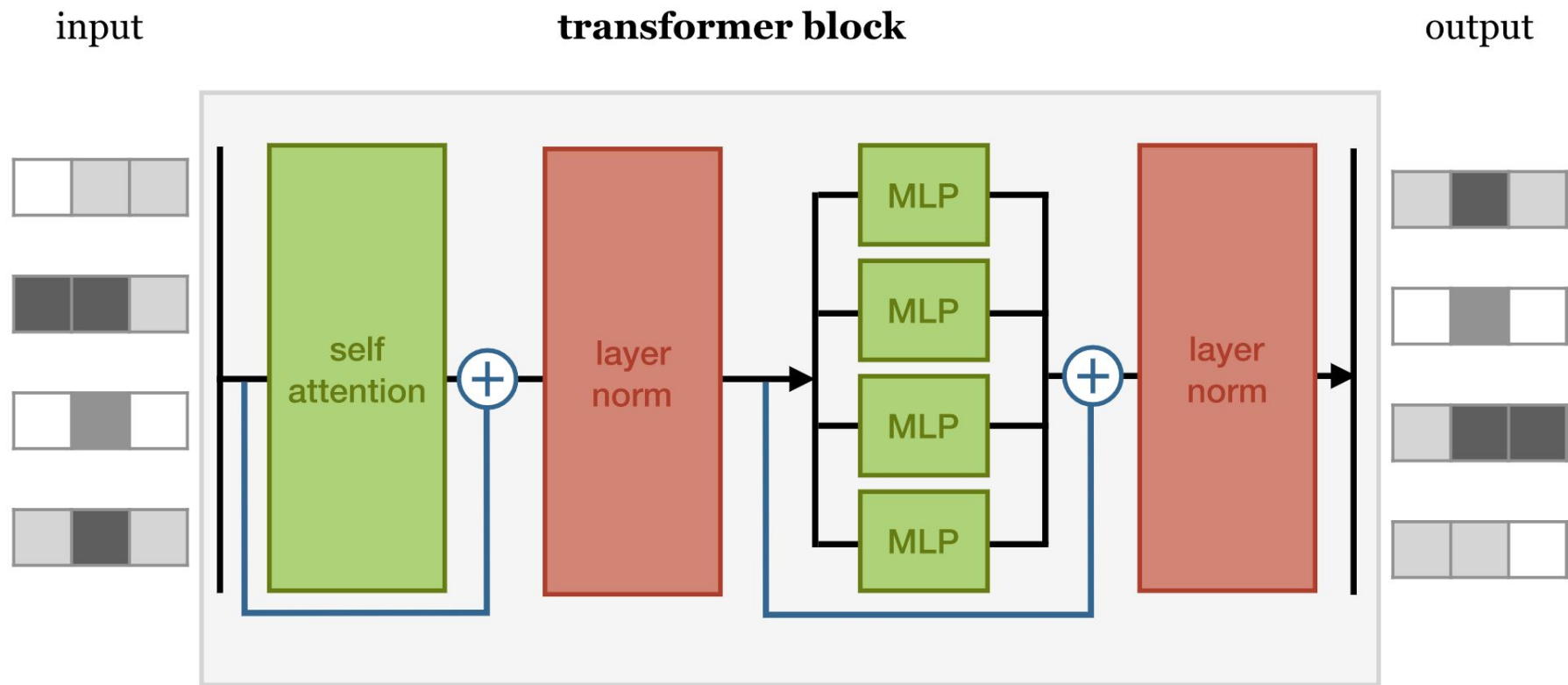
ФКН ПМИ, 3 курс

Николай Шаталов

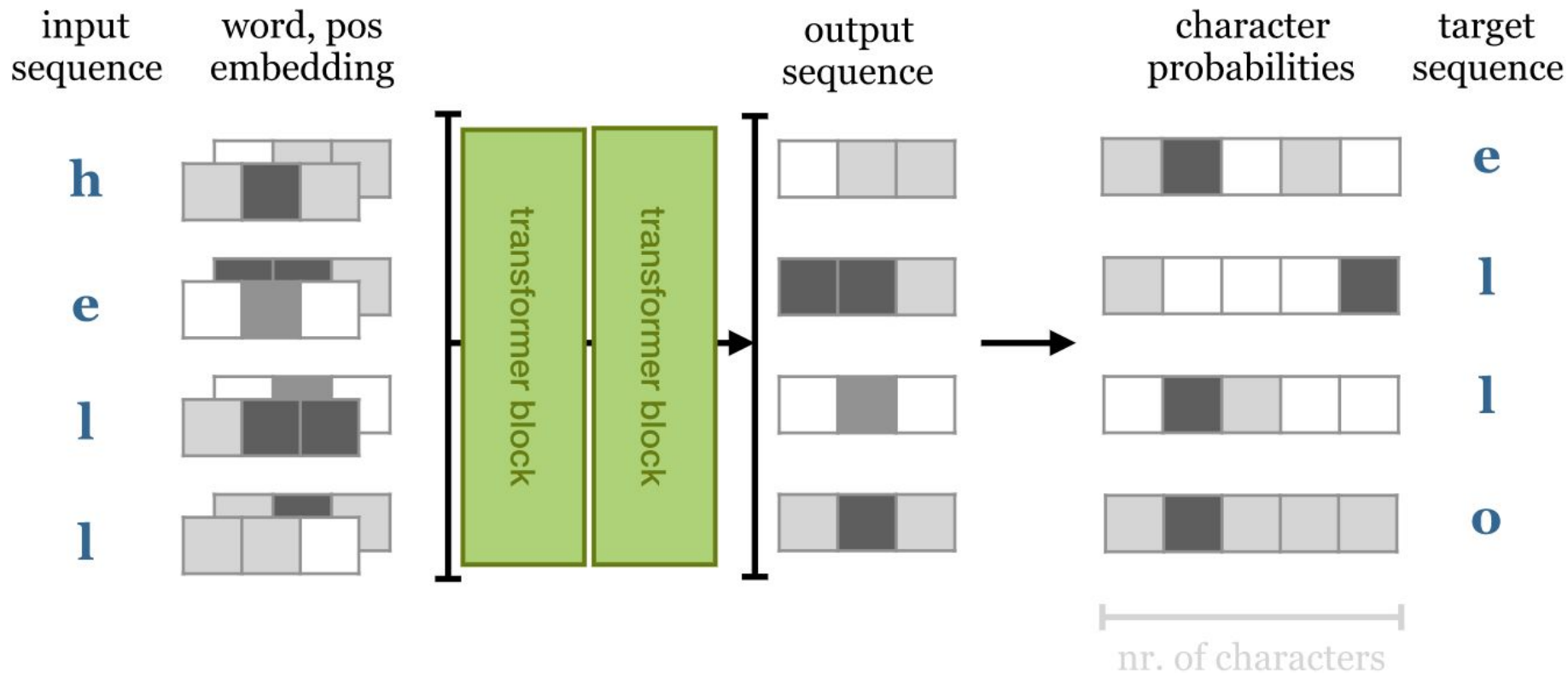
МГУ, 1 курс маг.

Вадим Филиппович,
Яндекс

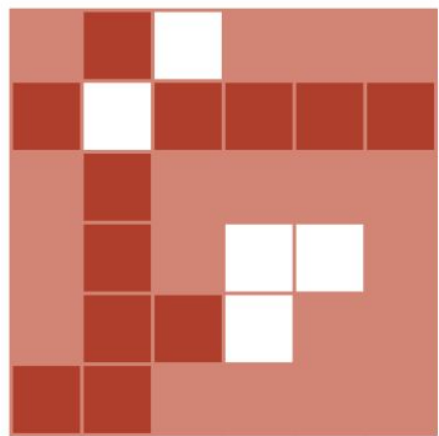
Впитываем как губка:
Базовый трансформер



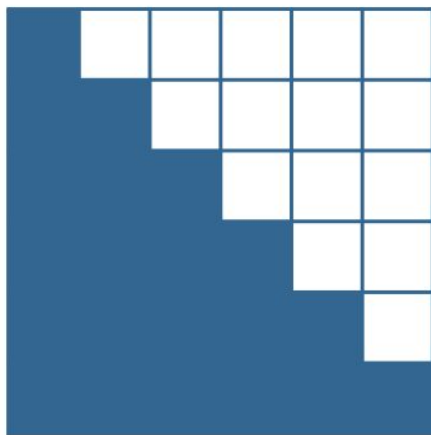
Источник: <http://www.peterbloem.nl/blog/transformers>



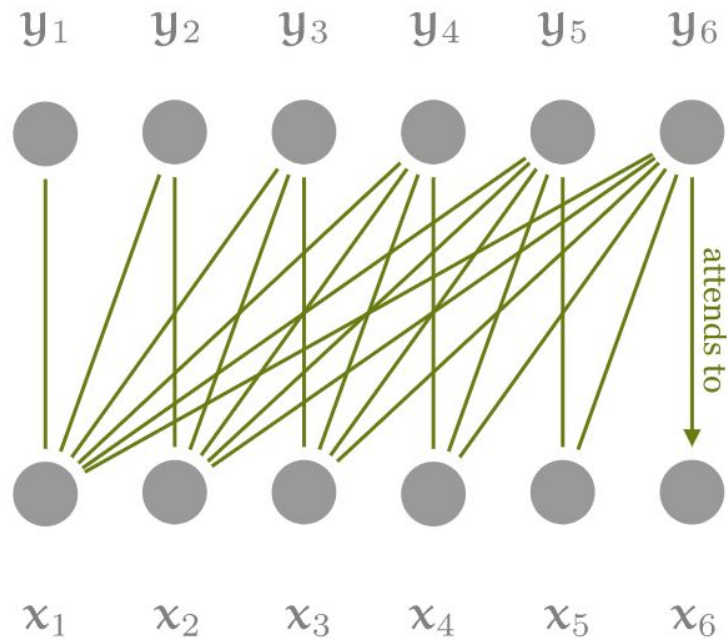
Источник: <http://www.peterbloem.nl/blog/transformers>



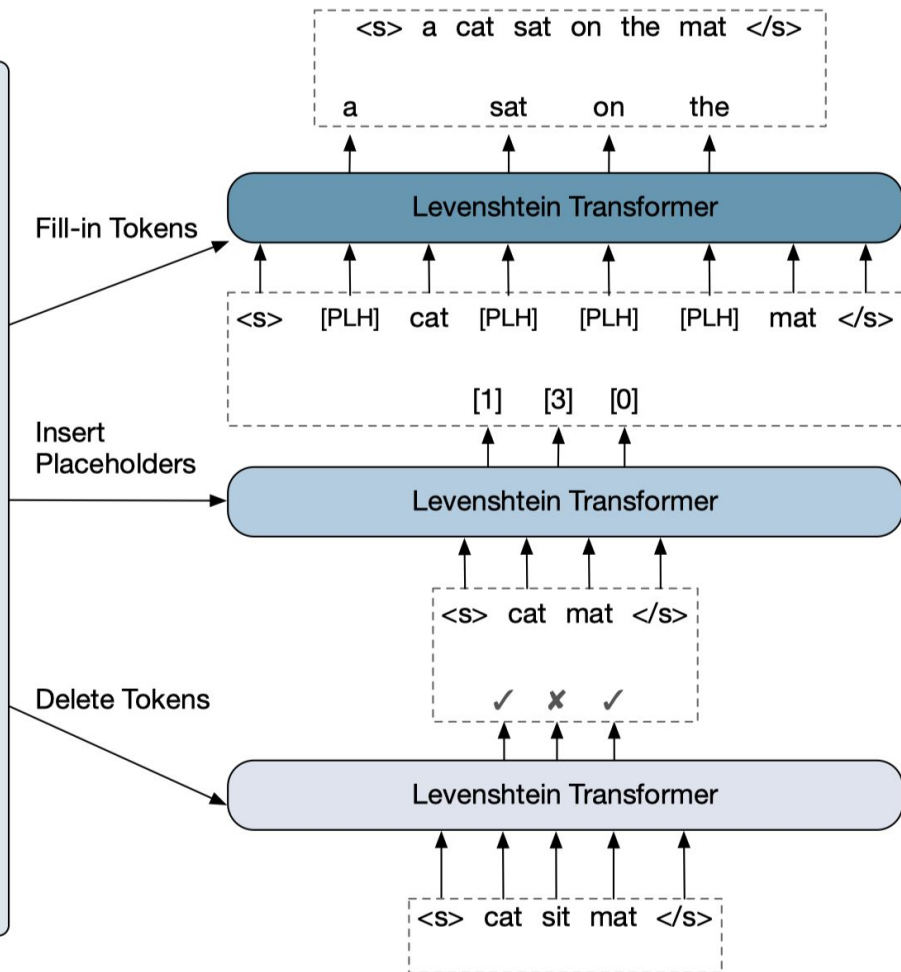
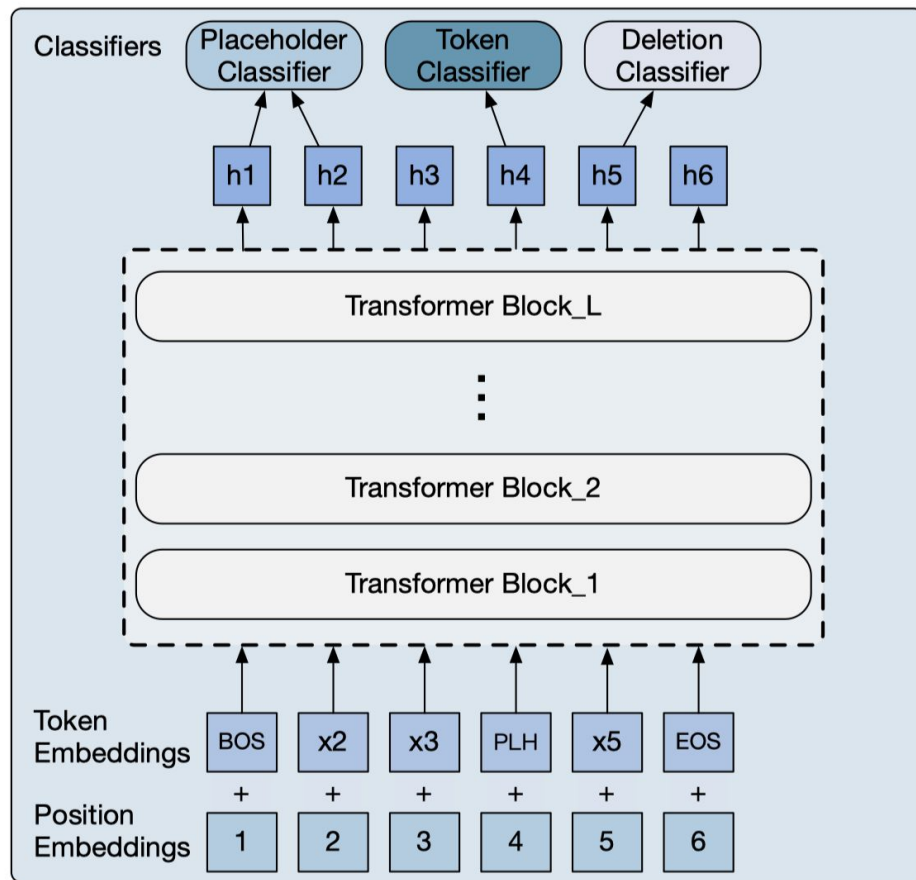
raw attention weights

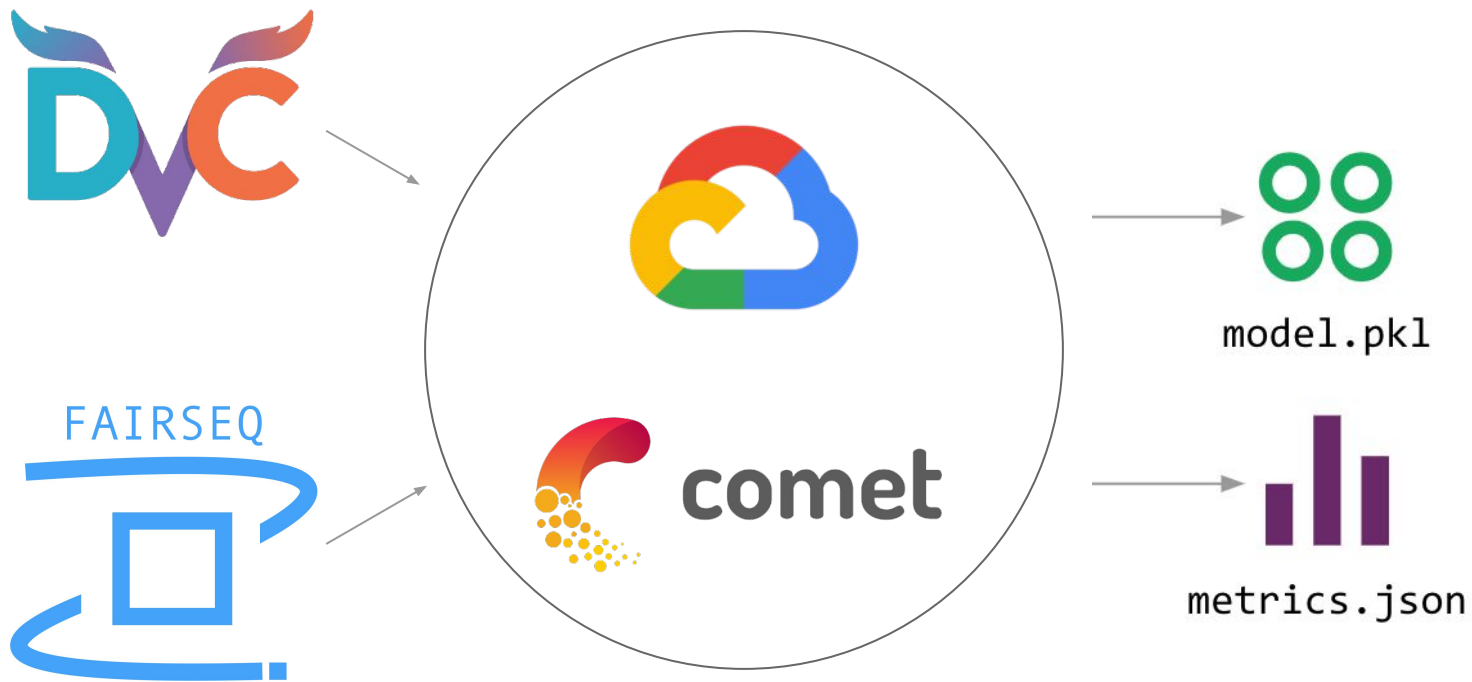


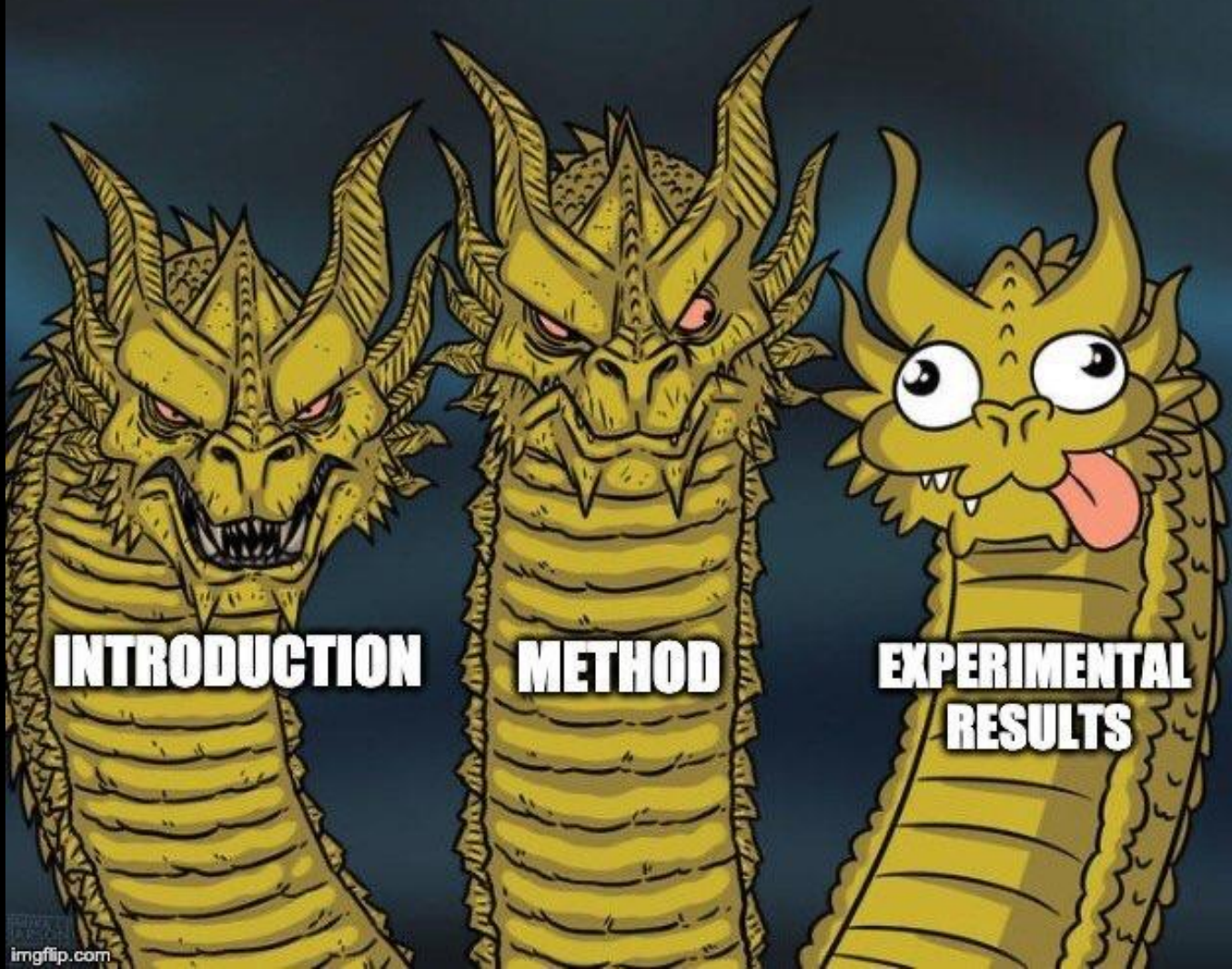
mask



Вырабатываем вкус: Трансформер Левенштайна







INTRODUCTION

METHOD

**EXPERIMENTAL
RESULTS**

Пример из Gigaword

Предобработанное исходное предложение:

**chairman of soccer australia david
hill resigned suddenly late thursday**

Желаемое обобщение:

australian soccer chief resigns

Наше предложение:

**australian soccer soccer
chairman resigns**

Steps

#	Result
1	<unk> <unk> <unk> <unk> <unk>
2	<unk>soccer <unk>soccer <unk>david <unk>chairman <unk>resigns
3	soccer soccer david chairman resigns
4	<unk> soccer <unk> <unk> resigns
5	<unk>australian soccer <unk>chairman <unk>chairman resigns

6 australian soccer chairman ~~chairman~~-resigns

7 australian soccer <unk> chairman resigns

8 australian soccer <unk> ~~soccer~~ chairman resigns

9 australian soccer ~~soccer~~ chairman resigns

10 australian soccer <unk> chairman resigns

11 australian soccer <unk> ~~soccer~~ chairman resigns

Меряем качество: **ROUGE-L**

- В основе — LCS
- Лишь условная оценка семантических совпадений

Трансформер Левенштайна: Версия 1

	Dataset	Metric	Transformer		Levenshtein Transformer	
			greedy	beam4	oracle	teacher
Quality ↑	Ro-En	BLEU	31.67	32.30	33.02	—
	En-De	BLEU	26.02	26.56	24.43	26.67
	En-Ja	BLEU	42.86	43.68	42.36	43.17
		ROUGE-1	34.91	35.19	35.57	36.08
	Gigaword	ROUGE-2	17.05	17.58	17.11	18.33
		ROUGE-L	32.66	32.98	33.55	33.81
Speed ↓	Ro-En	Latency (ms) / I_{DEC}	326 / 27.1	349 / 27.1	97 / 2.19	—
	En-De	Latency (ms) / I_{DEC}	343 / 28.1	369 / 28.1	126 / 2.88	92 / 2.05
	En-Ja	Latency (ms) / I_{DEC}	261 / 22.6	306 / 22.6	112 / 2.61	106 / 1.97
	Gigaword	Latency (ms) / I_{DEC}	116 / 10.1	149 / 10.1	98 / 2.32	84 / 1.73

Трансформер Левенштайна: **Версия 2**

	Dataset	Metric	Transformer		Levenshtein	Transformer
			greedy	beam4	oracle	distillation
Quality ↑	Ro-En	BLEU	31.67	32.30	33.02	33.26
	En-De	BLEU	26.89	27.17	25.20	27.27
	En-Ja	BLEU	42.86	43.68	42.36	43.17
		ROUGE-1	37.31	37.87	36.14	37.40
	Gigaword	ROUGE-2	18.10	18.92	17.14	18.33
		ROUGE-L	34.65	35.13	34.34	34.51
Speed ↓	Ro-En	Latency (ms) / I_{DEC}	326 / 27.1	349 / 27.1	97 / 2.19	90 / 2.03
	En-De	Latency (ms) / I_{DEC}	343 / 28.1	369 / 28.1	126 / 2.88	92 / 2.05
	En-Ja	Latency (ms) / I_{DEC}	261 / 22.6	306 / 22.6	112 / 2.61	106 / 1.97
	Gigaword	Latency (ms) / I_{DEC}	116 / 10.1	149 / 10.1	98 / 2.32	84 / 1.73

Три датасета:

РИА Новости: 1М статей на русском

Gigaword: 4М коротких новостей

CNN & DailyMail: 300К длинных

Инференс Левенштайна

ROUGE-L (Gigaword): **39.6-39.7**

SoTA (Gigaword): **36.87**

Средняя скорость (с/шт)

Левенштайн: **0.137**

Базовый + pre-train: **0.231**

Сравнение моделей на РИА

Исходная статья:

МОСКВА, 16 ноя - РИА Новости. Президент России Владимир Путин в ходе телефонного разговора поздравил президента Киргизии ...

МОСКВА, 16 ноя - РИА Новости. Президент России Владимир Путин в ходе телефонного разговора поздравил президента Киргизии Сооронбая Жээнбекова с днем рождения, сообщает пресс-служба Кремля.

"В ходе телефонного разговора президент Российской Федерации Владимир Путин тепло поздравил президента Киргизской Республики Сооронбая Жээнбекова с днем рождения", - говорится в сообщении.

Ранее сообщалось, что Путин направил коллеге поздравительную телеграмму, в которой отметил готовность к продолжению диалога и конструктивной работы.

Желаемое обобщение:

Путин провел телефонный разговор с президентом Киргизии

Наше предложение (Базовый):

путин поздравил президента киргизии с днем рождения

Наше предложение (Левенштайн):

путин поздравил с киргизии с днем рождения

Steps

#	Result
1	
2	<unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk> <unk>
3	<unk>путин <unk>поздравил <unk>с <unk>киргизии <unk>с <unk>с <unk>днем <unk>днем <unk>рождения
4	путин поздравил с киргизии е-е днем днем рождения

5 путин поздравил с киргизии <unk> днем рождения

6 путин поздравил с киргизии <unk>с днем рождения

7 путин поздравил с киргизии е-днем рождения

8 путин поздравил с киргизии <unk> днем рождения

9 путин поздравил с киргизии <unk>с днем рождения

Демо



phramer.sdll.io



Phramer

[Home](#)

[GitHub](#)

Levenshtein Demo

Please submit a short text for summarization. For better results, use news-like stories.

Source text

Start processing

Chairman of soccer Australia, David Hill, resigned suddenly late Thursday

Steps

- | # | Result |
|---|---|
| 1 | <unk> <unk> <unk> <unk> <unk> |
| 2 | <unk>soccer <unk>soccer <unk>david <unk>chairman
<unk>resigns |
| 3 | soccer soccer david chairman resigns |
| 4 | <unk> soccer <unk> <unk> resigns |