

STAP: Simultaneous Translation and Paraphrase



Hello!

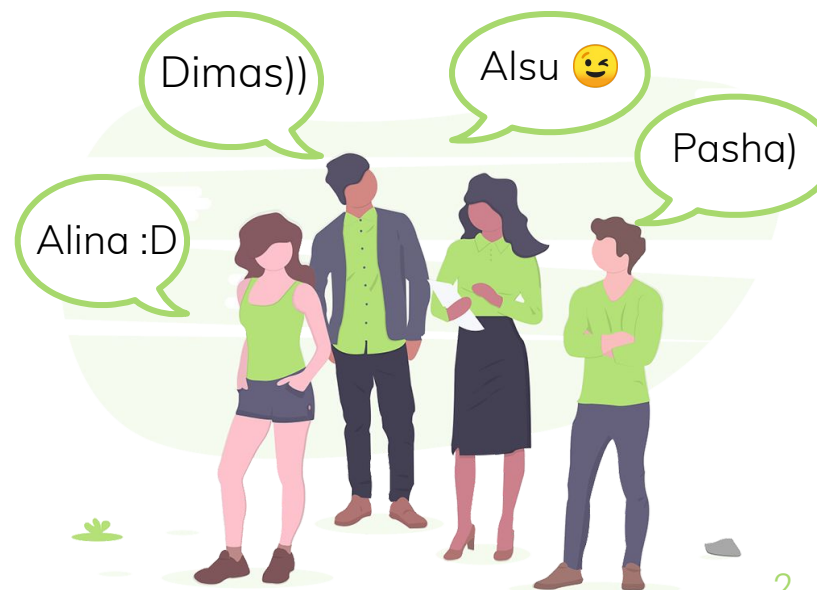
We are:

Alina Ilalova - @a3198i

Dmitry Grigorev - @gridmi

Alsu Vakhitova - @palette_knife25

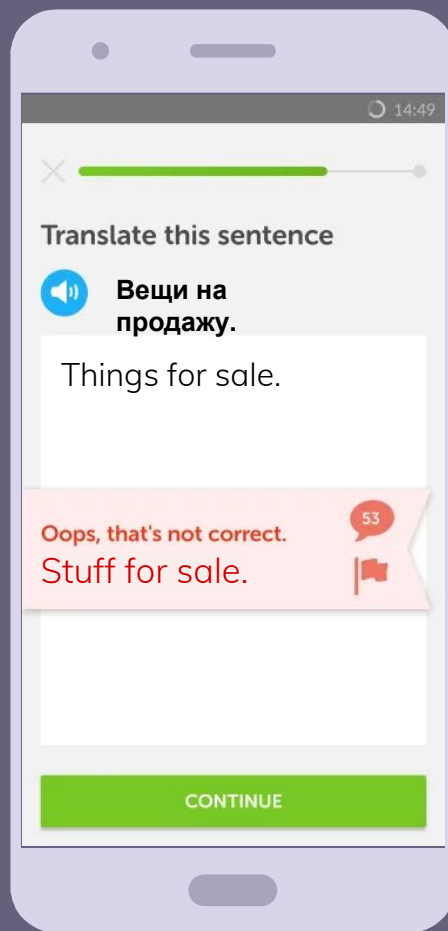
Pavel Kozlov - @KozlovPA



1. Problem

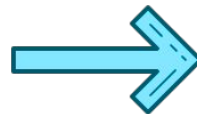


One-to-one

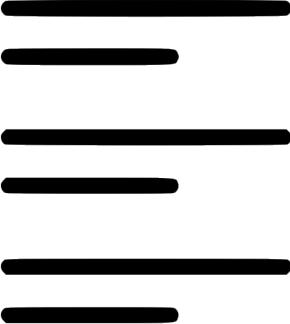


Our idea

Знаете, ведь
так мы и
умираем -
в одиночестве.



1
2
3

Three horizontal black lines are positioned to the right of each number, stacked vertically.

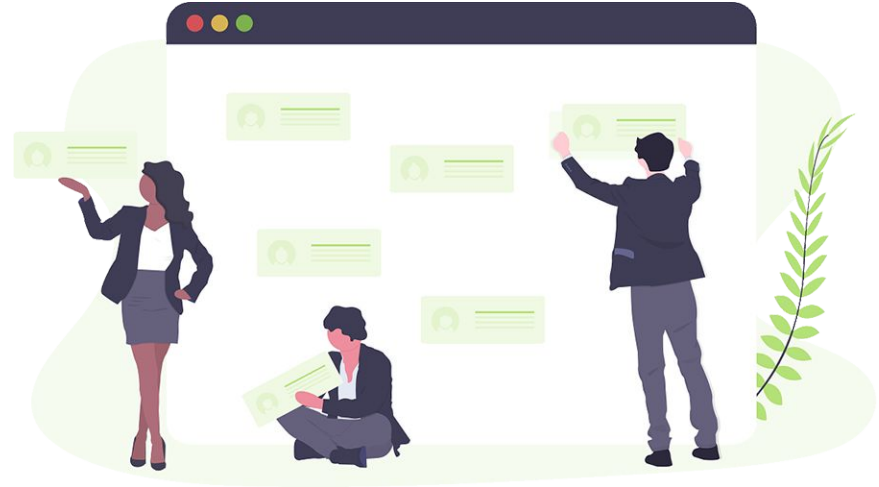
Potential users



new language learners

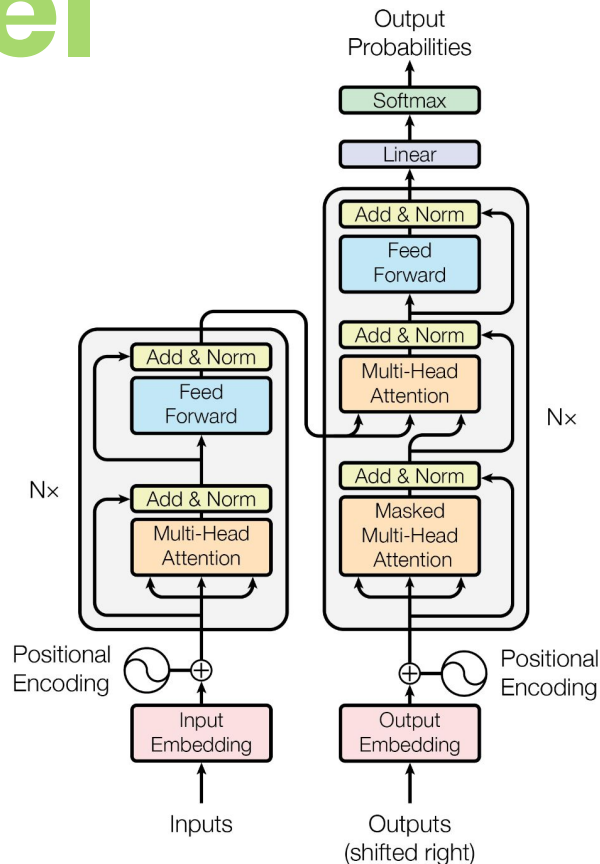
2.

Implementation details



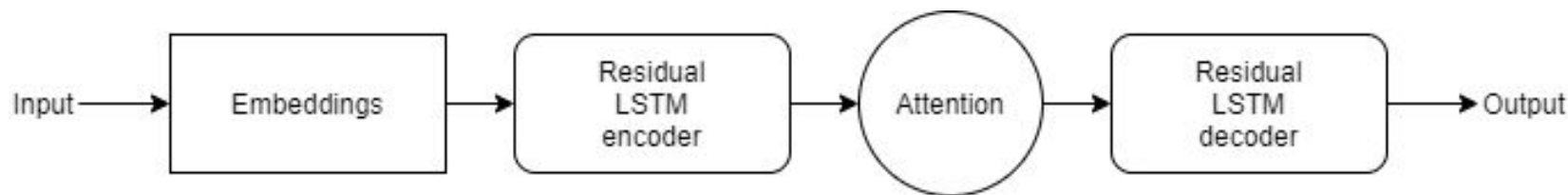
Translation model

pre-trained FairSeq
model by Facebook AI
Research



Paraphrase model

- Trained on MS COCO (also tried on WikiAnswers)
- Tensorflow
- 100 000 iterations (~3 hours)



MS COCO dataset



A man holding a child while standing at the fence of an elephant zoo enclosure.

...



The horse and puppy are separated by the mesh fence.

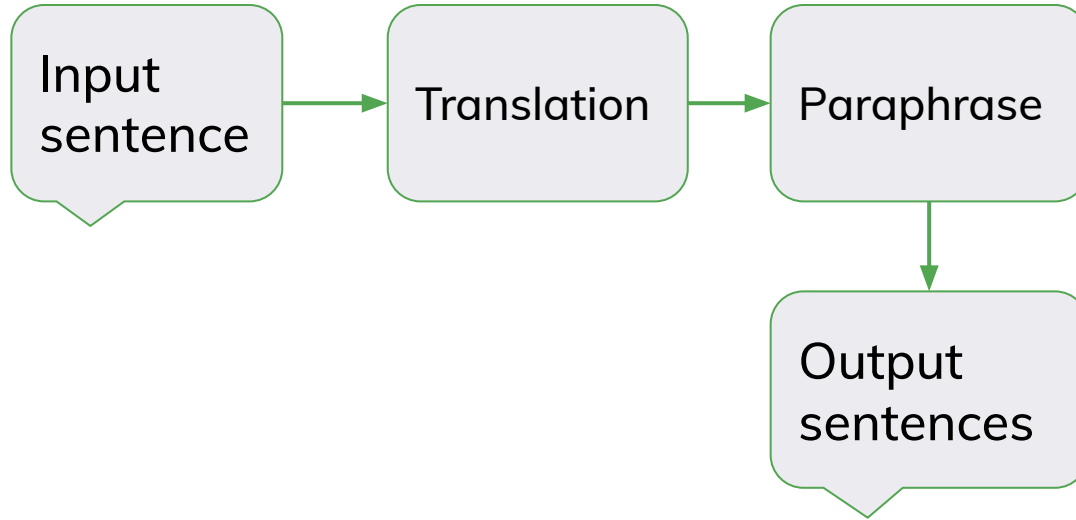
...

WikiAnswers dataset

clusters of questions tagged
by WikiAnswers users as
paraphrases

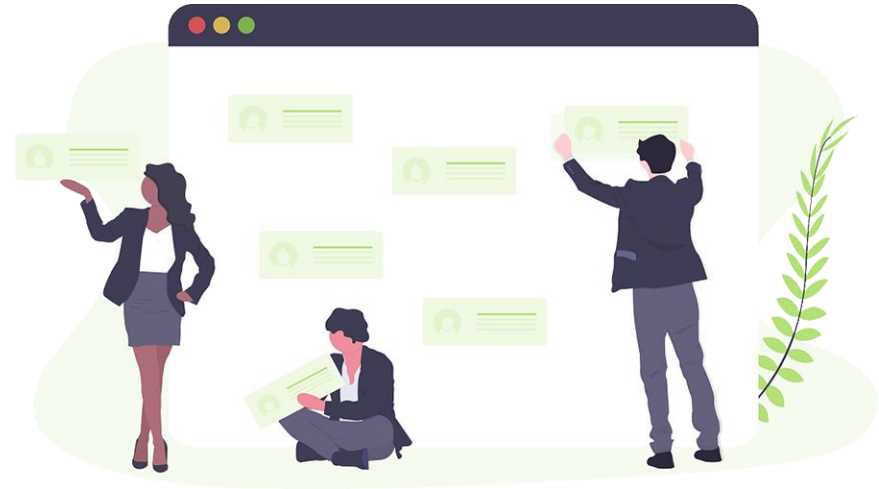
Cluster 1	
Q: <i>"Who was the Britain's leader during WWI?"</i>	temporal
Q: <i>"Who ran Britain during WWI?"</i>	
Q: <i>"Who was the leader of Britain during World War One?"</i>	
A: [https://en.wikipedia.org/wiki/h._h._asquith , https://en.wikipedia.org/wiki/david_lloyd_george]	
Cluster 2	
Q: <i>"largest city located along the Nile river?"</i>	comparison
Q: <i>"largest city by the Nile river?"</i>	
Q: <i>"What is the largest city in Africa that is on the banks of the Nile river?"</i>	A: [https://en.wikipedia.org/wiki/cairo]
Cluster 3	
Q: <i>"John Travolta and Jamie Lee Curtis acted in this film?"</i>	compositional
Q: <i>"Jamie Lee Curtis and John Travolta played together in this movie?"</i>	
Q: <i>"John Travolta and Jamie Lee Curtis were actors in this film?"</i>	A: [https://en.wikipedia.org/wiki/perfect_(film)]
Cluster 4	
Q: <i>"Who is the first human landed in Mars?"</i>	empty answer
Q: <i>"Who was the first human being on Mars?"</i>	
Q: <i>"first human in Mars?"</i>	A: []

Final workflow

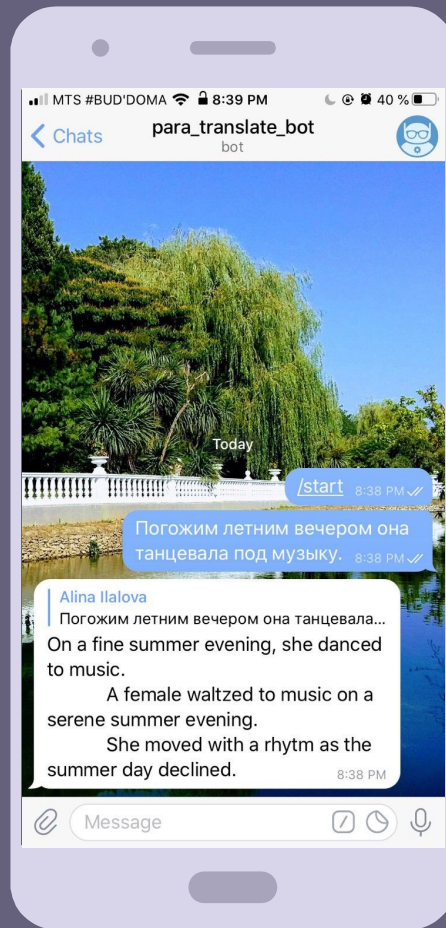


3.

Application Production Techniques



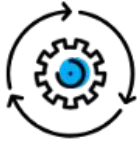
Telegram Bot



Continuous Integration Pipeline



Commit
change



Trigger
build



Build



Notify of
build
outcome




Run
tests



Notify of
test
outcome

Scrum Activities

Machine Translisation project  Updated 5 days ago

Filter cards + Add cards Fullscreen Menu

2 In progress

- Project Deliverable D2: Technical Report Assignment**
#12 opened by palette-knife25
Sprint 6
- Project Deliverable D3: Slides Assignment**
#13 opened by palette-knife25
Sprint 7

Automated as **In progress** Manage

12 Done

- Joining two models into one**
#14 opened by JiKook31
US-02 US-04 US-07
Sprint 6
- Prepare data for paraphrase model**
#10 opened by JiKook31
US-02
Sprint 5
- Progress submission**
#11 opened by palette-knife25
Sprint 5
- Implement paraphrase model**
#9 opened by JiKook31
US-02 US-07

Automated as **Done** Manage

7 User stories -> Tasks

- US-01 tasks:**
Data preprocessing
Create structure of the model
NMT experiment
Implement translation model
Added by palette-knife25
- US-02 tasks:**
Conduct paraphrase experiment
Implement paraphrase model
Prepare data for paraphrase model
Joining two models into one
Added by palette-knife25
- US-03 tasks:**
Added by palette-knife25

7 Project Backlog

- US-07**
Accuracy of the model
Description
As a Customer I want at least 70% accuracy and paraphrases are
Acceptance criteria
Given trained model
When evaluated on te
Then shows at least 7
Risks
The behaviour of the unpredictable and av
computational resour
enough for achieving
Priority (MoSCoW)
Could be done.
Estimation

Thanks!

Any questions?

