

# Headline generation: first sentence vs neural machine translation

Churikov N. S. (nikita@chur.ru),  
Sannikova E. (elena.sannikova59@gmail.com)

Saint Petersburg State University, Saint Petersburg, Russia  
Behavox, Saint Petersburg, Russia

In this article, we describe our experience of participating in a headline generation contest organized by VKontakte. We took the third place in the competition by modifying the baseline solution through the data cleaning. In addition, we tried to train and apply the transformer architecture combined with byte pair encoding, but this solution turned out to be worse than the baseline. At the end, we present our results on leaderboard for different solutions, and ROUGE scores on our test set.

**Key words:** text summarization, headline generation, Russian language

# Генерация заголовков: первое предложение против глубокого машинного перевода

Чуриков Н. С. (nikita@chur.ru),  
Санникова Е. (elena.sannikova59@gmail.com),

Санкт-Петербургский государственный Университет, Санкт-Петербург, Россия  
Behavox, Санкт-Петербург, Россия

В данной статье мы описываем наш опыт участия в конкурсе по генерации заголовков, организованном ВКонтакте. Мы заняли третье место в соревновании модифицировав пороговое решение через очистку от лишней информации. Это позволило нам перевалить за решение-baseline. Помимо этого, мы попробовали обучить и применить трансформер токенизируя данные через byte pair encoding, однако это решение оказалось хуже порогового. В конце мы приводим численные результаты, полученные в соревновании, а также оценки, полученные на нашей тестовой выборке.

**Ключевые слова:** автореферирование текстов, генерация заголовков, русский язык

## 1 Introduction

The problem of generating headers is a specific summarization problem, where output of a model is just a short single sentence. Conceptually, there are two approaches to the problem: extractive and abstractive summarization. In the case of extractive summarization, only the original text is used. Abstractive summarization is different in that the output text may contain words that were not in the source material.

In this article, we describe our results in the header generation competition, which was organized by VKontakte for Dialog Evaluation. In our work, we used extractive and abstractive summarization. The methods used by us are described in the experiments section, the results of our models on the leaderboard and on our test set are given in the results section. At the end of the work, we present our conclusions on the competition and our thoughts on what can be improved.

## 2 System description

In the headline generation competition, we’ve decided to compare classical summarization approaches (textrank), which do not require training and approaches based on neural networks. As a result, we’ve managed to get the results in the leaderboard only for part of our experiments, specifically for the first sentence, transformer-architecture [7] with bpe trained on Wikipedia [3] and trained on the competition data. We used the implementation of the transformer from the library of Open Neural Machine Translation [4]. The launch code for training, predictions, and models can be found in our repository <sup>1</sup>.

### 3 Data and training

Competition data was provided in the form of raw pieces of original html pages from Ria News. This means that there were various html tags and entities present in the data. And each news and a matching headline in raw text.

```
<p> <strong> <\strong> <\p> \n <p> <strong> Moscow, Dec 1 &nbsp; &mdash;
RIA news. <\strong> a fire in &nbsp; &nbsp; one of the &nbsp; &nbsp; workshops in
&nbsp; &nbsp; ...<\p>
```

As a result, the first thing we tried was to clean up the data from unnecessary information. So, we've made a preprocessor that removes all html tags and entities.

In addition, we’ve found that sometimes there is no text in the data, since the original news is represented by some form of image (for example, a screenshot of Twitter), and the news is purely Polish. These are all outliers we cleared data from.

Then, for training and validation of transformer, we've split the data in such a way that 90% of data goes for training and 10% for test.

## 4 Experiments

For all experiments, we used the preprobe as described above. Training models produced on 8 Nvidia K80.

### 4.1 First sentence

As described in [cite], the first sentence for automatically creating a news headline is a very strong baseline. In particular, and in the ongoing competition, the first offer of news was proposed by the threshold decision.

However, we used substantive knowledge about the data and as the first sentence we used the one in which there is no “ria news”.

## 4.2 Textrank

We used the classic extractive summarization algorithm - textrank [1]. To create a summary using keywords and original sentences we've set extraction size to be 20% of the original text.

<sup>1</sup>[https://github.com/kuparez/headline\\_generator](https://github.com/kuparez/headline_generator)

Algorithm\Score	1F	1P	1R	2F	2P	2R	LF	LP	LR
First sentence	0.23	0.16	0.44	0.10	0.07	0.21	0.16	0.15	0.40
Wiki BPE transformer (first sentence)	0.37	0.39	0.36	0.20	0.21	0.19	0.34	0.37	0.34
Wiki BPE transformer (first sentence, beam=5)	0.39	0.41	0.39	0.22	0.23	0.22	0.37	0.39	0.37
Wiki BPE transformer (first sentence, beam=1)	0.37	0.37	0.37	0.20	0.20	0.20	0.34	0.36	0.35
Ria BPE transformer (2000 first tokens)	0.36	0.37	0.35	0.18	0.20	0.18	0.33	0.36	0.33
Textrank summarization	0.14	0.09	0.41	0.05	0.03	0.17	0.09	0.08	0.38
Textrank keywords	0.09	0.07	0.18	0.00	0.00	0.01	0.05	0.05	0.14

Table 1: ROUGE-1,2,F1, precision and recall scores.

### 4.3 Transformer

Since the competition was held for only a month, we’ve decided to repeat the results described in the article VKontakte [2]. We took the implementation of the transformer from Open NMT, with the parameters described in the article.

Then, we tried to focus on what to use as input for the model. We tried the first sentence according to the rules above. We’ve also tried as an input first 2000 bpe tokens.

## 5 Results

As a result, none of the models we’ve trained could show better results than the first sentence. Unfortunately, we could not check some of our models on VK servers due to the restrictions on time complexity of our models. For this reason, we also present results on our test set.

Competition used mean of ROUGE-1,2,L F1 score.

## 6 Related work

The main article, that we’ve used as an inspiration, was the article VKontakte [2]. From there, we got the idea to use bpe [6], as well as which hyperparameters to use for training of transformer. Then, we’ve started to browse their citations and we’ve found that there were a some criticism about what to use as an input to abstractive headline generation algorithms. As a result, we stumbled upon an article where the choice of the first sentence as an input to models was being criticised and authors discussed ways to extract so-called Topic Sentence [5] is proposed.

Algorithm	Score
Baseline	0.19500071
First sentence	0.19502427
Wiki BPE transformer (first sentence, no beam search=5)	0.16397515
Ria BPE transformer (2000 first tokens, beam search=5)	0.16131584
Textrank summarization	0.10764881
Textrank keywords	0.06259589

Table 2: Results from evaluation by VK servers.

## 7 Conclusion and future work

We have seen that the first sentence is a very strong baseline in the problem of generating headers. We have tried various options for preprocessing data, but none of them helped us improve the results of transformer.

We think that in future work it is worth trying to generate a topic sentence on the 5W1H principle described in [5], and the problem may lie in the wrong choice of model hyper parameters.

## References

- [1] Federico Barrios, Federico López, Luis Argerich, and Rosa Wachenchauzer. Variations of the similarity function of textrank for automated summarization. *CoRR*, abs/1602.03606, 2016.
- [2] Daniil Gavrilov, Pavel Kalaidin, and Valentin Malykh. Self-attentive model for headline generation. In *Proceedings of the 41st European Conference on Information Retrieval*, 2019.
- [3] Benjamin Heinzerling and Michael Strube. BPEmb: Tokenization-free Pre-trained Subword Embeddings in 275 Languages. In Nicoletta Calzolari (Conference chair), Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018 2018. European Language Resources Association (ELRA).
- [4] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush. OpenNMT: Open-Source Toolkit for Neural Machine Translation. *ArXiv e-prints*.
- [5] Jan Wira Gotama Putra, Hayato Kobayashi, and Nobuyuki Shimizu. Incorporating topic sentence on neural news headline generation. 2018.
- [6] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. *CoRR*, abs/1508.07909, 2015.

- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017.