

# Cross-Language Document Extractive Summarization with Neural Sequence Model\*

*Захаров П.С., Пискун М.Г., Сельницкий И.С., Кваша П.А., Дьячков Е.А., Петров Е.Д.*

Московский физико-технический институт

В данной работе представлена модель образования краткого изложения текста на языке, отличном от языка документа. Для этого используется сокращение текста выбором предложений с последующим машинным переводом; при отборе предложений учитывается не только их содержание, но и оценка качества перевода. Исследуется зависимость качества сокращения от качества перевода. Перевод и сокращение осуществляются специально спроектированными для этих целей нейронными сетями. При этом базовая модель исследовалась на малом числе наборов данных; в этой работе идет дальнейшее рассмотрение переносимости этой модели на другие данные и внесение коррективов для улучшения модели в будущем.

**Ключевые слова:** *Сокращение текстов, машинный перевод, нейронные сети.*

## 1. Введение

Cross-language document summarization определяется как задача составления краткого изложения текста на языке, отличном от языка текста документа (Wan et al., 2010). В наше время огромный объем информации доступен в Интернете, однако большинство людей не может с ним ознакомиться из-за того, что информация представлена на незнакомом читателю языке. В этом исследовании мы фокусируемся на сокращении английского текста с последующим переводом на русский язык.

Существует не так много исследований, посвященных этой задаче по краткому изложению текста. Большинство из них полагаются на выбор предложений из переведенного текста. Например, Wan предложил ранжировать и выбирать предложения, учитывая два типа оценок уровня предложения: информативность и качество перевода [1]. Также Wan предложил совместно ранжировать предложения исходном языке и на языке, на который переводим с последующим выбором предложений согласно оценкам рейтинга [2].

## Список литературы

### Литература

- [1] Yong Zhang ; Jinzhi Liao ; Jiuyang Tang ; Weidong Xiao ; Yuheng Wang. Extractive document summarization based on hierarchical gru. In *2018 International Conference on Robots Intelligent System (ICRIS)*, pages 341– 346. IEEE, 2018.

[1]