

# Японский: буквенные n-граммы для распознавания

## Контроль НИР

Куликов А.В., гр. 397  
*Руководитель: Андрианов А.И.*

ABBYY-MIPT

Москва, 2017

- Japanese kanji/kana OCR.
- Существуют путающиеся символы, например:

お и ゐ

- *Цель работы:* построить и сравнить различные эвристики для исправления ошибок OCR, используя буквенную n-граммную модель японского языка.

- Buckets
- NGram
- Backoff (+Katz)
- Zipf

- Foundations of Statistical Natural Language Processing / C. D. Manning, H. Schutze.
- Efficient In-memory Data Structures for N-Grams Indexing / D. Robenek, J. Platos, V. Snasel
- Applying Conditional Random Fields to Japanese Morphological Analysis / T. Kudo, K. Yamamoto, Y. Matsumoto

# Спасибо