

# Анализ текстов на естественных языках

## Контрольная работа

**Задание 1.** Дан язык  $L = \{aa, bb, abab, baba, aaaa, bbbb\}$  над алфавитом  $\Sigma = \{a, b\}$ . Вычислить оценки максимального правдоподобия для параметров биграммной модели со сглаживанием по схеме “+1”.

**Задание 2.** Вычислить вероятность строки *abaaba* с помощью биграммной модели, построенной в задании 1.

**Задание 3.** Решается задача аннотирования гласных и согласных в словах переменной длины над алфавитом  $\Sigma = \{a, b, c\}$ , набор помет  $T = \{C, V\}$ . Обучающая выборка  $(\mathcal{X}, \mathcal{Y}) = \{(a, V), (b, C), (ab, VC), (ba, CV), (abb, VCC), (abba, VCCV)\}$ . Найти наиболее вероятную аннотацию для строки *bbaabb* в соответствии с построенной моделью.

**Задание 4.** Требуется построить систему автоматической расстановки переносов для русского языка на основе скрытой марковской модели. Выпишите спецификацию модели.

**Задание 5.** Дана вероятностная контекстно-свободная грамматика  $\langle N, \Sigma, R, S \rangle$ , где

- $N = \{S, A, B\}$
- $\Sigma = \{a, b\}$
- $R = \{S \rightarrow AB, A \rightarrow B, A \rightarrow a, B \rightarrow b\}$
- $q(S \rightarrow A) = 1.0, q(A \rightarrow B) = 0.1, q(A \rightarrow a) = 0.9, q(B \rightarrow b) = 1.0$

Найти наилучший (наиболее вероятный) разбор для строки *abb*