

Исследуемая открытая сеть массового обслуживания является сетью Джексона [1].

Исследована система массового обслуживания с двумя узлами, представляющая собой модель изолированного узла сети. Целью исследования было составить уравнение равновесия, найти стационарное распределение и проверить условия эргодичности.

Для такой сети было выведено уравнение трафика, учитывающее потоки входящих и исходящих сущностей, а также интенсивности обслуживания в каждом узле. Составлены уравнения глобального и локального равновесия. Найдено стационарное распределение, и для подтверждения корректности модели проведена проверка условий эргодичности [1, 2].

### Литература

1 Буриков, А. Д. Теория массового обслуживания: учебное пособие по спецкурсу / А. Д. Буриков, Ю. В. Малинковский, М. А. Маталыцкий. – Гродно : ГрГУ, 1984. – 108 с.

2 Малинковский, Ю. В. Теория вероятностей / Ю. В. Малинковский. – Минск : РИВШ, 2019. – 270 с.

**А. А. Чагочкин**

*(ГГУ имени Ф. Скорины, Гомель)*

## МЕТОД КЛАСТЕРНОГО АНАЛИЗА СВЯЗАННЫХ ТЕКСТОВ РАЗНЫХ ФОРМАЛЬНЫХ ГРАММАТИК НА ОСНОВЕ ТОКЕНОВ

Рассматривается задача анализа исходного кода проекта, включающего связанные тексты разных формальных грамматик: C#, JavaScript, Razor<sup>1</sup>, CSS.

Токен – лексема, базовая единица кода: операторы, ключевые слова, идентификаторы, константы и переменные.

---

<sup>1</sup> Объединение грамматик в Razor можно сравнить с научными публикациями, использующими формальные описания проблематики, доказательств и выводов, объединенных естественным языком.

Кластерный анализ на основе токенов позволяет группировать лексически связанные тексты и строить графы семантических связей на основе повторяющихся элементов. Используются принципы методов векторизации и кластеризации текстов на естественном языке (TF-IDF): общий словарь токенов, взвешивание по частотной характеристике и по цитируемости в совокупности текстов. Метод использует индексы вместо символьных представлений.

Кластеризация файлов проводится на основе векторов наличия токенов (1 есть, 0 нет). Критерии выбора векторов токенов: удельный вес в совокупности текстов (частота), коэффициент цитируемости между документами и количество токенов в векторе.

Словарь сортируется по частоте по убыванию. Используются три вектора с коэффициентами цитирования: от 0.9 (в 90% файлов) и 0.4 и 0.01. Размер векторов соответствует 0.001, 0.001 и 0.002 размера словаря соответственно и зависит от масштаба проекта. Увеличение вектора приводит к увеличению количества кластеров, что может соответствовать переобучению в управляемом обучении.

Оптимальное количество кластеров вычисляется методом Silhouette с косинусной метрикой. Кластеризация производится методом K-средних (K-means).

Метод демонстрирует эффективность при незначительных затратах вычислительных ресурсов на неразмеченных данных.

### Литература

1 Саммут, К. Энциклопедия машинного обучения. – Нью-Йорк : Спрингер, 2011. – 294 с.