

Отзыв научного руководителя д. ф.-м. н., профессора Д.Н. Бабина

На диссертацию Петюшко Александра Александровича «Биграммные языки», представленную на соискание ученой степени кандидата физико-математических наук по специальности 01.01.09 – дискретная математика и математическая кибернетика.

Еще в начале 20 века выдающимся русским ученым А.А. Марковым был создан аппарат цепей, впоследствии названных цепями Маркова. В дальнейшем этот аппарат получил широкое применение для распознавания и статистического моделирования для различных приложений. Обработка текстов на естественном языке, распознавание одномерных сигналов и речи, оптическое распознавание текстов, предсказание погоды – вот немногий перечень приложений цепей Маркова.

Как правило, в приложениях последовательность данных аппроксимируется вероятностями переходов цепи Маркова, а затем полученные вероятности переходов используются для предсказания дальнейшего поведения этой последовательности. Аналогом вероятностей переходов являются частоты появления пар соседних элементов в изучаемой последовательности. Эти частоты образуют матрицу биграмм.

Перед диссертантом была поставлена новая и обратная к традиционной задача: описать формальные языки последовательностей, имеющих одну и ту же матрицу биграмм, и получить классификацию матриц, соответствующую классификации формальных языков Хомского.

Следует заметить, что, несмотря на кажущуюся простоту постановки вопроса, для решения этой задачи автору зачастую приходилось привлекать сложные методы дискретной математики и комбинаторики. Александр Александрович качественно и всесторонне исследовал поставленные перед ним проблемы, и ему удалось успешно решить все поставленные задачи.

В первой главе диссертации введены основные понятия и показаны некоторые свойства матрицы кратностей биграмм, а также получены критерии непустоты, конечности и счетности биграммных языков.

Во второй главе, получена точная формула количества слов конечного биграммного языка. Такие результаты редко встречается для формальных языков. Аналогичные результаты были использованы для оценки числа различных последовательностей ДНК.

Автор рассмотрел вопрос о доле матриц, которые задают пустые, конечные или счётные биграммные языки; также вопрос о доле матриц задающих регулярные, контекстно-свободные или контекстно-зависимые биграммные языки.

В третьей главе изучены так называемые “биграммные языки с закольцовыванием”, в которых к набору биграмм добавляется биграмма, состоящая из последней и первой буквы рассматриваемого слова. В работе построена их классификация.

Также в третьей главе показано, что рассмотренные выше задачи о мощности и классификации для n -граммных языков при $n > 2$ сводятся к соответствующим задачам для биграммных языков.

Фактически, Александр Александрович установил критерий того, что матрица биграмм задаёт конкретный тип формального языка из иерархии Хомского и получил интерпретацию этой иерархии в терминах эйлеровых циклов графа матрицы биграмм. Получилось, что тип языка зависит от числа степеней свободы в структуре циклов указанного графа. Для этого понадобилось определить и исследовать новую функциональную систему графов с операциями сложения и вычитания, принципиально не совпадающей с векторным пространством. В этой функциональной системе были исследованы вопросы единственности разложения графов на элементарные циклы.

Если посмотреть на матрицу n -грамм, как на хэш-функцию, то n -граммный язык есть ничто иное как язык коллизий этой хэш-функции. Из

результатов автора следует, что альтернативная попытка задать контекстно-зависимый язык коллизий через грамматики имеет большую вычислительную сложность. Это фактически закрывает задачу использования n-граммных хэш-функций в базах данных большого размера.

Считаю, что Петюшко А.А. проделана большая, важная и актуальная работа. Полученные результаты хорошо оформлены, изложение сопровождается большим числом примеров, иллюстрирующих доказываемые свойства. Автор демонстрирует высокий уровень научных способностей и огромные творческие возможности. В целом, работа Петюшко А.А. имеет большое теоретическое и практическое значение, и является серьезным вкладом в теорию формальных языков.

Считаю, что рассматриваемая диссертация представляет собой законченное научное исследование. Результаты получены лично автором и прошли всестороннюю квалифицированную апробацию, они опубликованы в ведущих научных журналах.

Сказанное позволяет заключить, что данная диссертационная работа целиком и полностью удовлетворяет требованиям ВАК к диссертациям на соискание ученой степени кандидата физико-математических наук по специальности 01.01.09 – дискретная математика и математическая кибернетика, а её автор – Петюшко Александр Александрович – заслуживает присуждения ему ученой степени кандидата физико-математических наук.

Научный руководитель
д.ф.-м.н., профессор



Бабин Д.Н.

Подпись Д.Н. Бабина удостоверено
И.о. декана
механико-математического факультета
МГУ им. М.В. Ломоносова, профессор



Чубариков В.Н.