

ОТЗЫВ НАУЧНОГО КОНСУЛЬТАНТА

о диссертационном исследовании Ерболовой Асель Серикановны
«Исследование структур естественного языка в задаче идентификации ботов»,
представленном на соискание степени доктора философии (PhD)
по образовательной программе «8D06101 – Информатика, вычислительная техника и
управление»

Диссертационная работа Ерболовой Асель Серикановны представляет собой комплексное исследование, которое раскрывает важные аспекты функционирования естественного языка как сложной системы. В диссертационной работе подробно рассматриваются структурные особенности языка с применением методов математического и статистического анализа, что позволяет выявить внутренние размерности языковых фрактальных структур и их топологические характеристики. Особое внимание уделено разработке и обоснованию эффективных методов идентификации текстов, созданных людьми, и текстов, сгенерированных автоматизированными системами (ботами), что имеет важное значение для современных задач обработки и анализа естественного языка. Работа охватывает широкий спектр задач, включая построение векторных представлений (эмбедингов) слов и n-грамм, анализ семантических траекторий, применение методов персистентной гомологии и топологического анализа данных, а также разработку и экспериментальную проверку классификационных моделей для различения текстов, написанных людьми и ботами, на материале различных языков.

Основные научные результаты, полученные соискателем и выносимые на защиту, заключаются в следующем:

1. На материале корпусов национальной литературы для широкого круга языков экспериментально подтверждена гипотеза о естественном языке как о самоорганизующейся критической системе. Показана универсальность степенных распределений для языковых единиц. Важным результатом является выявление исключительного статуса эсперанто, демонстрирующего гауссово, а не степенное распределение, что подтверждает его искусственную природу.

2. Разработан и применён метод анализа «семантических траекторий» текстов, позволивший количественно оценить их хаотичность через расчёт энтропии и

сложности. Кластерный анализ на основе этих метрик выявил типологические сходства и различия между языками.

3. Предложен комплекс методов для оценки внутренней (фрактальной) размерности языковых структур (для n -грамм $n=1,2,3$). Результаты указывают на мультифрактальную природу языка и инвариантность размерностных характеристик относительно способа построения векторных представлений, что свидетельствует об устойчивости выявленных закономерностей.

4. Методы топологического анализа данных (персистентная гомология) были адаптированы для изучения крупномасштабной структуры семантического пространства языка. Это позволило выявить и количественно описать наличие «дыр» (топологических особенностей) в пространстве n -грамм, которые являются свойством языка в целом, а не артефактом конкретной выборки текстов.

5. На основе выявленных фундаментальных характеристик разработана и успешно протестирована классификационная модель для решения прикладной задачи идентификации текстов, написанных человеком, и текстов, сгенерированных ботами. Ключевой особенностью методологии является строгое разделение выборки по принципу «незнакомого бота» на тесте, что обеспечивает высокую достоверность результатов. Модель демонстрирует эффективность свыше 96% на материале нескольких языков.

В целом представленная диссертационная работа является завершённым научным трудом, в котором построен научно обоснованный алгоритм и предложены методы анализа крупномасштабной структуры естественного языка, основанные на статистических, геометрических и топологических подходах. Предложенные методы позволяют эффективно выявлять внутренние размерности языковых фрактальных структур и обнаруживать топологические особенности, существенно расширяя понимание динамики и организации языковых систем. Кроме того, разработаны и экспериментально проверены классификационные модели, способные с высокой точностью различать тексты, созданные людьми, и тексты, сгенерированные ботами, что имеет важное практическое значение для современных систем обработки естественного языка.

Данная научная работа является результатом самостоятельного комплексного исследования докторанта, обладает актуальностью и научной новизной, что позволяет считать Ерболову Асель Серикановну достойной соискания степени доктора философии (PhD). Исходя из изложенного, считаю, что диссертационная работа Ерболовой А. С. на тему «Исследование структур естественного языка в задаче идентификации ботов» является завершённой, соответствует установленным требованиям и может быть рекомендована к защите.

Научный консультант

д.ф.-м.н., профессор НИУ ВШЭ

Департамента анализа данных и

искусственного интеллекта _____

Громов В.А.

Подпись Громова В.А. удостоверяю

