

**Индекс УДК 004.89** Прикладные системы искусственного интеллекта.  
Интеллектуальные системы, обладающие знаниями

Методы интеллектуального анализа естественного языка для разработки системы  
искусственного интеллекта в здравоохранении

Гречищева С. Р.  
(СПб НИУ ИТМО, Санкт-Петербург)  
Научный руководитель – Ковальчук С.В.  
(СПб НИУ ИТМО, Санкт-Петербург)

**Введение.** Базы данных медицинских информационных систем содержат в себе ценные эмпирические знания о процессах лечения и закономерностях протекания заболеваний. Данные знания возможно использовать для разработки интеллектуальных информационных систем, к которым возможно отнести системы искусственного интеллекта, характеризующиеся наличием баз знаний, наличие моделей на основе баз знаний, интерпретируемостью, способностью формировать решения на основе слабоструктурированных, неполных данных. Электронные медицинские записи (ЭМЗ) характеризуются слабой структурой и разнородностью информации по характеру и по форме представления. Подробная информация может быть использована для разработки прогностических моделей. Качество и структурированность данных являются одним из основных вопросов, когда мы имеем дело с разработкой моделей на данных, а также баз знаний, их обеспечивающих. Более того, в различных контекстах информация из ЭМЗ может интерпретироваться различным образом. В этой связи, разработка методов анализа слабоструктурированных данных ЭМЗ с учетом специфики предметной области здравоохранения является значимой научной задачей.

**Цель работы:** разработка метода анализа слабоструктурированных данных электронных медицинских записей для обеспечения задач разработки компонентов интеллектуальных информационных медицинских систем.

**Базовые положения исследования:**

- 1) выполнить аналитический обзор предметной области интеллектуальных методов анализа слабоструктурированных данных: базовая структура методов интеллектуального анализа, основные подходы к интеллектуальному анализу;
- 2) разработать словарь синонимов и аббревиатур для сокращения размерности текстовых векторов, описывающих эпизоды лечения;
- 3) разработать метод синтаксического анализа данных, нацеленный на уменьшение размерности текстового вектора, описывающий эпизод лечения пациента;
- 4) провести вычислительный эксперимент для факторизации показателей состояния пациентов с сердечно-сосудистыми заболеваниями для обоснования разработанного метода.

**Промежуточные результаты:** разработан метод анализа слабоструктурированных данных электронных медицинских записей для повышения качества моделей на данных в задачах персонализированной медицины Р4; выполнен аналитический обзор предметной области интеллектуальных методов анализа слабоструктурированных данных ЭМК; разработан метод структурирования данных и извлечения информации из ЭМЗ; разработан словарь синонимов и аббревиатур для сокращения размерности текстовых векторов, описывающих эпизоды лечения; проведен вычислительный эксперимент для факторизации показателей состояния пациентов с сердечно-сосудистыми заболеваниями для обоснования разработанного метода; проведен эксперимент демонстрирующий уменьшение текстового вектора.