

РАЗРАБОТКА ВИРТУАЛЬНОГО ДИАЛОГОВОГО ПОМОЩНИКА ДЛЯ ПОДДЕРЖКИ ПРОВЕДЕНИЯ ДИСТАНЦИОННОГО ЭКЗАМЕНА НА ОСНОВЕ ПОНИМАНИЯ ЕСТЕСТВЕННОГО И МАТЕМАТИЧЕСКОГО ЯЗЫКА

Матвеев А.Ю.

(Университет ИТМО)

Научный руководитель – д. т. н. Матвеев Ю.Н.

(Университет ИТМО)

В работе описывается разработка автоматизированной системы для проведения письменного экзамена. Рассматриваются вопросы понимания текста, в том числе математических символов и выражений, машиной — выявления смысловой и эмоциональной составляющей, генерации осмысленных вопросов к тексту и определения корректности ответа, данного человеком.

Исследования выполнены за счет финансирования университета ИТМО в рамках НИР № 620183 «Разработка виртуального диалогового помощника для поддержки проведения дистанционного экзамена на основе моделей-трансформаторов и понимания естественного и математического языка».

В современных информационных системах недостаточно развиты и применяются технологии разговорного искусственного интеллекта для человеко-машинной коммуникации, что ограничивает их гибкость и масштабируемость. В работе рассматриваются подходы к построению системы, способной анализировать текст как на естественном языке, так и текст с математическими символами и выражениями, генерировать вопросы — в том числе с математическими символами, математическими выражениями из оригинального текста, а также модифицированными математическими выражениями, и анализировать и оценивать ответы пользователей системы, включая интерпретацию математических символов и выражений, введенных пользователем.

В области машинного обучения и обработки естественного языка по отношению к математическим символам и выражениям выделяются несколько задач, основными из которых являются сопоставление текста на естественном языке формальной математической модели и сопоставление математических символов и выражений терминам или выражениям на естественном языке. В работе основное внимание уделяется второй задаче. Основной сложностью данной задачи является то, что зачастую одни и те же математические символы, в зависимости от контекста, могут отражать различные концепции и выступать в роли синонимов к различным терминам, а математические выражения, отражающие определенные концепции — формулы, теоремы и т.п., могут быть записаны различными символами и отличаться по форме записи.

В работе рассматриваются два основных подхода к решению задачи. Один подход состоит в интерпретации математических символов и выражений как расширения естественного языка и применении к ним стандартных методов обработки естественного языка. Другой подход состоит в интерпретации математических символов и выражений как обособленных сущностей и применении к ним систем правил, специального подготовленных для выделения расширенных, по сравнению с естественным языком, наборов признаков. По отношению к таким наборам признаков далее применяются статистические методы, например кластеризация и классификация, которые позволяют построить статистические модели. Такие модели, далее, могут применяться для расширения онтологий, построенных для оригинальных текстов, и в дальнейшем могут применяться для генерации вопросов или оценки корректности высказываний.