Практический сравнительный анализ методов распознавания именованных сущностей для цифровых сервисов оияи

А. Ильинаa\*

a Объединенный институт ядерных исследований, ул. Жолио-Кюри, д. 6, г. Дубна, Московская область, 141980, Россия

\*e-mail: *annailina@jinr.ru*

**Аннотация** – В современных условиях стремительного роста объемов текстовой информации эффективное извлечение именованных сущностей становится одним из ключевых аспектов анализа данных в различных областях науки и техники. Задача анализа текстовых данных крайне актуальна для ряда внутренних цифровых сервисов Объединенного института ядерных исследований (ОИЯИ), в частности, в контексте развития системы планирования и учета экскурсий JINRex, что определяет необходимость выбора наиболее эффективных методов обработки информации. Проведено сравнение методов извлечения именованных сущностей в алгоритме обработки текстовых данных об экскурсиях. Результаты работы алгоритма позволяют получать статистическую информацию о проводимых мероприятиях из текстов на естественном языке, содержащих краткое описание названия экскурсии и ее целевой аудитории. В данной работе представлен практический сравнительный анализ различных подходов к выделению именованных сущностей с акцентом на их применимость при решении задач анализа и обработки текстов в рамках поддержки и развития внутренних цифровых сервисов ОИЯИ.