

Разработка методов оценки качества проектов умных решений при их внедрении в умном городе

И. М. Холмогоров (Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург)

Научный руководитель: С. А. Митягин (Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург)

На данный момент отбор умных решений производится на основе методики и регламента критериальной оценки, ранжирования и отбора проектов при их включении в состав концепции “Умного Санкт-Петербурга”. Все предлагаемые проекты, после заполнения анкеты на сайте “Умного Санкт-Петербурга”, попадают в общий реестр “умных решений”. Затем каждый из них проходит длительную процедуру конкурсного отбора. Между тем, строго структурированная форма заявок открывает возможность использования методов области компьютерной лингвистики для создания многопараметрической системы поддержки принятия решений, что упростит модерацию заявок и оптимизирует процесс их рассмотрения.

Цель работы: разработать концепцию автоматической оценки качества проектных заявок умных решений.

Исследование базируется на методах машинного перевода, информационного поиска и машинного понимания естественного языка. На данный момент разработана концепция алгоритма автоматической оценки качества проектов умных решений при их внедрении в умном городе.

Согласно разработанной концепции, алгоритм автоматической оценки проектов умных решений осуществляется в три шага, а именно: оценка полноты заполнения заявки; оценка правописания; оценка связности. Первые два этапа представляют собой достаточно тривиальную задачу ввиду существования доступных и эффективных программных инструментов, в то время как смысловой анализ текста на предмет непротиворечивости требует применение методов из области компьютерной лингвистики. Для анализа текста заявки на связность и отсутствие противоречий предложены два основных процесса: процесс извлечения и преобразования информации в упрощенные модели; процесс анализа полученных моделей. Задача первого состоит в извлечении из текстов заявок ключевой информации о значимых объектах, параметрах, отношении между ними и событиях планируемых, или уже имеющих место быть. Сам процесс состоит из нескольких этапов: сегментация текста (выделение в тексте предложений словоформ), морфологический анализ, синтаксический анализ. В результате текст будет преобразован в упрощенные модели, представляющие из себя плоские таблицы, где наименования колонок – концептуализационные элементы (субъект, действие, объект), а номеру каждого ряда соответствует порядковый номер самих предложений. Задача второго процесса заключается в анализе полученных моделей на предмет их связности. Это достигается посредством применения программного инструмента Word2vec: полученные данные преобразуются в вектора N-мерного пространства; при этом расстояния между ними зависят от их семантической схожести, таким образом ближайшими “соседями” выступят синонимы и различные тэги, присущие каждому конкретному концептуализационному элементу. Последующий анализ расстояния между геометрическими эквивалентами разных пунктов заявки позволит оценить их смысловую близость или выявить противоречия. Иными словами, заявка, цель которой состоит в озеленении городского пространства, но в описании которой нет

соответствующих слов и словосочетаний, характерных данной тематики (дерево, куст, пруд, животные, тропа и т.д.) получит низкую оценку по критерию связности. Данному анализу будут подвержены следующие пары пунктов заявки: “Цель проекта” и “Описание проекта”; “Описание проекта” и “План проекта”; “Сфера применения проекта” и “Описание проекта”; “Сфера применения проекта” и “Показатели эффективности проекта”; “Сфера применения проекта” и “Проблемы города, которые позволяет решить этот проект”. Аналогичным образом будут оценены пункты, описывающие пользу проекта для той или иной категории граждан, путем сравнения с специально составленным облаком тэгов: “Опишите пользу проекта для жителей города”; “Опишите пользу проекта для маломобильных граждан”; “Опишите пользу проекта для граждан с детьми”; “Опишите пользу проекта для льготных категорий граждан”; “Опишите пользу проекта для малого/среднего бизнеса”.