

Акционерное общество «Крибрум»
(АО «Крибрум»)
ИНН 7731661104 КПП 773101001
ОГРН 1107746952433

Юр. адрес: территория инновационного центра
«Сколково», Большой бульвар, д. 42, стр. 1, эт.
0, пом. 263, р.м. 20, г. Москва, 121205
Почтовый адрес: ул. Верейская, д. 29, стр.134,
г. Москва, 121357
+ 7 499 390 67 13, www.kribrum.ru

ФГУП «ГРЧЦ»
Врио генерального директора
Нестеренко Р.В.

23.06.2022 № 22/06-23.01
№ _____ Дата _____

Уважаемый Руслан Васильевич!

ФГУП «ГРЧЦ» 25 апреля 2022 года отправило нам для ознакомления техническое задание (ТЗ) на создание системы «Вепрь», предназначенной для выявления точек социальной напряженности в публикациях в СМИ и социальных сетях.

Наши специалисты внимательно изучили данное ТЗ. Наша компания более 10 лет занимается сбором и анализом информации социальных сетей и решением задач, описанных в полученном ТЗ (поиск и анализе деструктивного контента в социальных сетях, блогах, форумах и СМИ, анализ и прогнозирование информационных кампаний). Всё это время мы сотрудничаем со спецслужбами России по этим задачам. Поставленные в ТЗ задачи понятны, мы имеем практический опыт решения их в «боевой» обстановке.

Понимая национальную важность разработки, описанной в ТЗ, считаем своим долгом предупредить о возможных проблемах с реализацией проекта при условии выполнения его в точности по данному техническому заданию.

По оценке специалистов нашей компании, имеющаяся постановка задачи обладает рядом следующих проблемных мест:

1. Техническое задание – чрезвычайно детальное, в нём на крайне низкоуровневом техническом языке описаны требования, конкретная техническая схема реализации и архитектура проектируемой системы, как в целом, так и отдельных элементов.

То есть, вместо установления общих требований к исполнителю по целям, задачам, функциональному наполнению и критериям качества проектируемой системы в данном ТЗ

Оригинал *М*

Вх. 64389 от 23.06.22

ФГУП «ГРЧЦ»	
Вх. №	65121
от	27 ИЮН 2022



предоставить систему с заранее заданной архитектурой и набором модулей заданного строения, вплоть до самых мелких. Нам представляется, что это не позволяет найти и использовать ни существующие на рынке системы, ни отдельные готовые компоненты, даже если бы они у кого-то существовали.

Фактически, такое детальное ТЗ с заранее предписанной архитектурой и принципами работы потребует разработки системы и всех её компонентов «с нуля».

2. Мы на основании нашего опыта оцениваем такую разработку «с нуля» примерно в 50-100 человеко-лет, что делает практически невозможным исполнение ТЗ в заявленный срок 12 месяцев.
3. Кроме того, есть серьёзные сомнения, что даже реализованная в точности по данному ТЗ система покажет реальные требуемые результаты и уровень качества. Это риск неудачи проекта возникает по следующим причинам:

3.1. Разработанная в ТЗ детальная архитектура системы и её отдельных модулей – чисто теоретическая, не проверенная на практике. По нашему опыту, существующие системы, работающие с реальными данными из социальных сетей и СМИ, устроены совершенно иначе. Фактически, разработка системы «с нуля» будет одновременно проверкой умозрительной инженерной гипотезы, описанной в ТЗ – с возможным «отрицательным научным результатом».

3.2. Например, общий подход обсуждаемого ТЗ к построению системы таков: шесть анализаторов текста (выделение именованных сущностей, распознавание тональности и т.п.) выстроены в единый «конвейер» обработки данных, так что каждый из анализаторов последовательно принимает на вход результат работы предыдущего анализатора.

Это красивая конструкция, привлекательная «на бумаге», но при таком подходе будет практически гарантированно происходить перемножение ошибок предыдущего и следующего этапов, что даст кардинальное ухудшение параметров качества системы «на выходе».

Даже если предположить, что все анализаторы сделаны качественно и имеют достаточные точность 90% и полноту 60%, то расположив их последовательно, как требует ТЗ, в конце цепочки получим точность 0.53 (0.9^6) и полноту 0.05 (0.6^6). Анализом с такими низкими итоговыми параметрами качества при реальном

применении системы на «грязном Интернете» нельзя будет предметно воспользоваться.

Нужно заметить, что на практике некоторые из анализаторов, требуемых по ТЗ, скорее всего, будут иметь значительно более низкие параметры полноты/точности, что еще больше снизит итоговую полноту и точность Системы в целом;

3.3. В ТЗ подробно описано, что является объектом поиска системы. А именно, высказыванием, которое может нести социальную напряженность, считается фраза, с одной стороны, направленная на определенное лицо или организацию (более широко именуемые «Сущность»), с другой – имеющая определенную (вероятно, отрицательную) тональность. Этот подход оставляет за пределами рассмотрения большой объем публикаций, в которых объект высказывания менее конкретен (не лицо или организация, а социальная группа, явление, государство в целом и т. д.) или вовсе не назван (скрыт за анафорами, эллипсисами, иносказаниями и т.д.). Это еще более снизит параметры качества системы, в область рассмотрения будет попадать весьма узкий класс потенциально важных сообщений.

Подводя итог, можно предположить, что выполнить в обсуждаемом ТЗ поставленную задачу получится, если привлечь с рынка участников (участника) с готовыми работающими системами выкачки и анализа текстовой и графической информации, добившись выполнения от них требований проекта по функциональности и качеству, но при этом ослабив требования к точному исполнению именно той архитектуры системы, которая до мельчайших деталей предписана в ТЗ и которой пока не существует в природе.

Мы готовы подробно обсудить описанные выше вопросы и проблемы на специальном совещании, если руководство ГРЧЦ сочтёт необходимым такое совещание созвать.

С уважением,

Генеральный директор АО «Крибрум»



/Л.Г. Лемэр/