Отчет по проекту Газпром Нефти.

Нами были разобраны презентации отчетов, представленные Газпром нефтью. На основе данных презентаций была построена семантическая модель МТР «Кабельная продукция». Семантическая модель представлена в прилагаемой презентации на слайдах 1-4. Также были проанализированы источники, используемые в представленных презентациях на доступность.

На основе построения семантической модели и анализа источников, используемых в эталонных материалах было выявлено, что большая часть источников являются платными источниками, либо результатами исследований маркетинговых и консалтинговых агентств у которых Газпром нефть заказывал исследования. Часть источников является отчетами Газпром Нефти и компаний поставщиков кабеля, к которым мы скорее всего тоже не сможем получить полный доступ. Полный список источников, к которым у нас нет доступа, представлен на слайде 6. Так как у нас нет доступа к этим источникам, то задача достигнуть 80% точности на основе их эталонных презентаций труднодостижимая задача.

При условии, что мы будем использовать только открытые источники, результаты могут отличаться от тех, что представлены в эталонных материалах. Поэтому мы предлагаем пересмотреть критерий оценки итогового результата.

На данном этапе работы нами был спроектирован и реализован модуль сбора текстовых данных из открытых источников. В качестве поиска релевантных документов использовались поисковые системы Google и Яндекс, чтобы получить как можно большую и точную выборку документов. На слайде 5 в презентации показа диаграмма классов данного модуля. В дальнейшем мы будем расширять нашу архитектуру при добавлении новых модулей.