
Система анализа существенных условий документов

Введение

Проведение госзакупок и тендеров на сегодняшний день, динамично развивающийся технологический процесс с постоянно меняющимися требованиями, возникающими в связи с изменением правовых документов. С другой стороны количество госзакупок и тендеров измеряется сотнями тысяч, тк все госзакупки и основная часть закупок крупных корпораций идет чз этот механизм.

Количество электронных площадок госзакупок по 44-фз - 8 шт, по 223-фз - 34 шт, количество крупных электронных площадок для коммерческих закупок и тендеров больше 100, а всего коммерческих электронных площадок более 5000. Участие в закупках и тендерах сопровождается регистрацией и подачей необходимых документов. Это ставит перед организаторами торгов ресурсоемкую задачу проверки поданных документов на подлинность, на соответствие нормативным документам, на соответствие заявленным ценам и тд.

Проведение закупки по тендерным правилам весьма непростое и дорогостоящее мероприятие в силу того, что

1. от организаторов требуется организовать прозрачный процесс
2. от участников требуется в подготовить и предоставить кучу документов в определенном формате
3. от организатора затем требуется все пакеты документов от всех участников обработать и проверить

Проблемы, связанные с проверкой документов

В силу того, что документы, хотя и предоставляются в электронном виде, это не означает, что они могут быть легко обработаны автоматически. По сути, сегодня электронные документы используются как замена бумажным носителям, при этом их обработка как проводилась вручную человеком, так и проводится. Факт наличия человека в процессе обработки и анализа критически важных документов приводит

к появлению ошибок при такой обработке. Кроме того, остается еще и вопрос как контролировать качество при такой ручной обработке, и кто будет контролировать контролеров.

Существуют следующие проблемы:

- как автоматически извлекать из неструктурированных документов, в которых информация не занесена в именованные поля, структурированную фактологическую информацию (например, название компании, ее адрес, сумму оплаты и т.п.)
- как извлекать структурированную фактологическую информацию, когда для ее описания могут быть использованы различные языковые конструкции, синонимы и т.п.
- как извлекать смысловую/логическую информацию из таких документов (например условия предоставления технической поддержки, условия оплаты и т.п.)
- как контролировать качество извлечения информации

Обзор текущих решений

На сегодняшний момент готовых решений в данной области нет. Смежными технологиями, которые могут быть использованы в решении являются следующие:

1. Технологии анализа тем (topic analysis), в частности латентный семантический анализ (LSA)
2. Семантические графы/тензоры
3. Технологии обработки запросов на естественном языке (NLP)
4. Технологии семантического моделирования

ЦЕЛИ

Задачи смыслового анализа документов в контексте проведения закупок можно разделить на следующие группы:

1. Выявление существенных условий в контексте данной закупки, то есть определенных как существенные в документации по проведению данной закупки

-
2. Выявление существенных условий с точки зрения законодательства
 3. Анализ на соответствие условий от поставщика условиям проведения закупки
 4. Логический и смысловой (семантический) анализ на предмет соответствия эвристическим правилам (например, компания поставщик похожа в некотором смысле на фирму-однодневку)

РЕШЕНИЕ

Предлагаемое решение будет основано на следующей концепции:

1. На основании правила проведения закупки специалистом строится семантическая модель, содержащая логические правила. Это делается в полуавтоматическом режиме.
2. Закупочные документы от участников загружаются в систему и по ним строятся семантический граф/тензор.
3. Проводится автоматическая проверка закупочных документов на предмет соответствия требованиям закупки, описанным в семантической модели закупки и регламенту проведения закупок.
4. В процессе проверки автоматически извлекается фактологическая информация.
5. Кроме того, в полуавтоматическом режиме проводится извлечение логической информации из текстов документов, что позволяет повысить качество в режиме обучения с учителем.
6. Полуавтоматический анализ хода разбора позволит дорабатывать эвристические модели/правила в системе.

Описание проекта

Продукт будет представлять из себя рабочее место бизнес аналитика и/или специалиста организации, реализующее следующий функционал:

- задание семантической модели правил проведения закупки/закупок
- система проверки/верификации модели закупки
- загрузка и анализ закупочных документов от участников

-
- извлечение структурированной информации в БД для последующей обработки
 - среда для извлечения логической информации из текстов, её анализа, и дообучения системы
 - средства постанализа и доработки эвристических логических моделей
 - открытые API для интеграции с СУБД и СЭД

Полезные ссылки:

1. <http://zakupki.gov.ru/epz/main/public/home.html>
2. <https://open-torg.ru/>
3. Руководство по регистрации организаций и пользователей в Единой Информационной Системе закупок.

<http://zakupki.gov.ru/epz/main/public/document/view.html?searchstring=§ionid=1086>

4. <https://zakupkihelp.ru/uchastniku-zakupok/elektronnye-torgovye-ploshadki.html#part-1>
5. <https://sbis.ru/tenders>