

Семантический анализ некоторых типов предложений Правил дорожного движения

М. И. Менькин¹

Данная работа относится к области автоматической обработки юридических документов и посвящена семантическому анализу Правил дорожного движения. Описаны основные этапы семантического разбора предложений типа «Уступить дорогу» и «Запрещено». Используя синтаксические шаблоны, мы ограничиваем поиск и реализуем синтаксический и семантический анализ этих предложений.

Ключевые слова: семантический анализ, синтаксический анализ, юридический документ, правила дорожного движения.

1. Введение

Существует задача автоматической обработки текстов юридических документов. По сравнению с произвольным текстом на естественном языке (ЕЯ) структура текста юридического документа более-менее определена, т.е. существует регламент по оформлению таких документов, есть конкретный набор используемых лексем с определённым значением, есть регулярность в синтаксисе однотипных предложений и т.д. А среди юридических документов есть такие, которые по своей сути являются алгоритмом, записанным на ЕЯ и формализованным с юридической точки зрения. В качестве примеров здесь можно привести документы по бухгалтерскому учёту (семантическим анализом которых занимался Евгений Перпер [1]) или документ, касающийся правил дорожного движения — Постановление Правительства РФ «О правилах дорожного движения» (ПДД) [2].

Одна из наших главных задач заключается в построении формальной модели юридического документа [2]. Актуальность работы можно сформулировать следующим образом:

- Если есть формальная модель юридического документа, то появляется возможность отвечать на вопросы, сформулированные на ЕЯ.

¹ Менькин Михаил Игоревич — м.н.с. каф. математической теории интеллектуальных систем мех.-мат. ф-та МГУ, e-mail: mihail_mn@inbox.ru.

Menkin Mikhail Igorevich — junior researcher, Lomonosov Moscow State University, Faculty of Mechanics and Mathematics, Chair of Mathematical Theory of Intellectual Systems.

Вопросы будут переводиться на язык этой же формальной модели. Для ПДД такой формальной моделью является модель дорожных ситуаций [3], [4], а примерами вопросов могут являться вопросы из экзамена на знание ПДД.

- Если есть формальная модель юридического документа, то мы сможем выполнить проверку текста юридического документа на непротиворечивость. Также, если мы захотим внести изменения в существующий текст или добавить новые предложения, то можно будет проверить противоречат ли внесённые изменения старому тексту или нет. Такая ситуация возникает когда выходит новая редакция юридического документа. Для ПДД проверка на непротиворечивость заключается в разрешении новых дорожных ситуаций без противоречий между разными пунктами текста документа.

Стоит отметить, что юридические тексты с алгоритмом внутри себя должны пониматься одинаково каждым читающим, т.е. семантика документа не должна зависеть от читающей интеллектуальной системы. Это используется для разработки модели семантики документа (первый этап семантического анализа документа) и, в дальнейшем, алгоритма «текст документа» → «модель семантики документа» (второй этап семантического анализа).

Данная работа продолжает исследование, начатое в [3], [4].

2. Типы предложений и синтаксические шаблоны

Относительно поставленной задачи можно считать, что все предложения в тексте [2] разбиваются на 3 класса: предложения–определения (необходимы для формирования модели семантики закона, например пункт 1.4. «На дорогах установлено правостороннее движение транспортных средств»), предложения–правила (собственно правила, записанные на ЕЯ, например пункт 17.3. «При выезде из жилой зоны водители должны уступить дорогу другим участникам движения»), мусорные предложения (предложения, удаление которых из текста не изменяет семантику предложений первых двух классов, например, такой текст из пункта 1.1.1: «абзац исключен. - Постановление Правительства РФ от 12.11.2012 N 1156»).

Такие классы предложений есть в любом содержательном документе, имеющим внутри себя алгоритмическую составляющую.

Для разбора предложений используются синтаксические шаблоны. Под *синтаксическим шаблоном* мы понимаем синтаксический граф, где каждая вершина представляет собой четвёрку «лемма», «синтаксическая

роль», «семантический класс», «ярус вершины». Наличие в синтаксическом шаблоне яруса и семантического класса избыточно, но служит для дополнительной проверки.

Возьмём пример из [4]. Предложения с манёвром «Уступить дорогу» имеют один из следующих синтаксических шаблонов (+ означает смежность вершин):

- (должен / должны, сказ., MODAL, 0) + (уступить, прям.доп., TO_CEDE, 1) + (дорогу, прям.доп., ROAD, 2);
- (обязан, сказ., TO_OBLIGE, 0) + (уступить, прям.доп., TO_CEDE, 1) + (дорогу, прям.доп., ROAD, 2);
- (имеет, сказ., EXISTENCE_AND_POSSESSION, 0) + (преимущество, прям.доп., CH_BENEFIT, 1).

Помимо этих синтаксических шаблонов, для предложений с манёвром «Уступить дорогу» есть наборы синтаксических шаблонов «КТО УСТУПАЕТ», «КОМУ УСТУПИТЬ», «СИТУАЦИЯ» [4]. Для предложений типа «Запрещено» есть наборы «СИТУАЦИЯ» (совпадает с предыдущим) и «МАНЁВР».

3. Процедуры алгоритма семантического анализа

Автоматическая обработка касается только предложений–правил. Выбираем те предложения, которые содержат ключевые слова «Уступить» и «Запрещено», т.к. среди предложений–правил их много.

Процедуры алгоритма семантического анализа:

- 1) Ищем предложения, содержащие ключевые слова, т.е. те, которые необходимо присутствуют в синтаксических шаблонах (в нашем случае «уступить» и «запрещено»); проверяем правильность синтаксической роли по соответствующему синтаксическому шаблону;
- 2) По каждому предложению из п.1 проверяем наличие каждого синтаксического шаблона;
- 3) По синтаксическому шаблону, который отвечает за манёвр, определяем вид и количество шаблонов правил (см. [4]);
- 4) По синтаксическим шаблонам «СИТУАЦИЯ» определяем свойства участков дорог в шаблоне правила;
- 5) Для предложений типа «Уступить»: по синтаксическим шаблонам «КТО УСТУПАЕТ» и «КОМУ УСТУПИТЬ» определяем расположение и свойства транспортных средств.

4. Заключение

На текущий момент создано около 80% синтаксических шаблонов для предложений типа «Уступить дорогу» и «Запрещено». Для нескольких простых предложений получены шаблоны правил. Получение шаблонов правил по предложениям, которые полностью разобраны по синтаксическим шаблонам, осложняется неоднозначностью отображений вершин графов, т.к. в шаблоне правила всегда 9 или менее вершин, в то время как в графе дороги их количество не ограничивается.

Автор выражает благодарность Гасанову Эльяру Эльдаровичу за постановку задачи. Для проведения синтаксического разбора предложений в работе использовались технологии NLP компании АБВУУ [5], [6], любезно предоставленные Дмитрием Черноусом.

Список литературы

- [1] Перпер Е.М., Гасанов Э.Э., Кудрявцев В.Б., “О семантическом анализе юридических текстов”, *Интеллектуальные системы*, **22:3** (2018), 45–88.
- [2] “Постановление Правительства РФ от 23.10.1993 N 1090 (ред. от 04.12.2018) «О правилах дорожного движения»”, *Консультант-Плюс*.
- [3] Менькин М.И., “Объектная модель правил дорожного движения”, *Интеллектуальные системы*, **23:1** (2019), 39–53.
- [4] Менькин М.И., “Автоматический перевод правил дорожного движения в теоретико-графовую формальную модель”, *Интеллектуальные системы*, **23:3** (2019), 7–31.
- [5] “АБВУУ InfoExtractor Demo”, <http://parsertest.abbyy.com/ru/>.
- [6] “Онтология Linguistic (АБВУУ)”, http://comprenoservice.abbyy.com/help/ru/ontologies_linguistic_linguistic.html.

Semantic analysis of some types of sentences of Road Traffic Law Menkin M.I.

The present paper is all about legal document processing and deals with semantic analysis of Road Traffic Law. We describe the basic procedures for semantic analysis of «Give way» and «Prohibitive»

types of sentences. We circumscribe search and implement parsing and semantic analysis of these sentences using syntactic patterns.

Keywords: semantic analysis, syntactic analysis, legal document, rules of the road.

References

- [1] Perper E.M., Gasanov E.E., Kudryavtsev V.B., “On the semantic analysis of juridical documents”, *Intelligent systems*, **22**:3 (2018), 45–88 (In Russian).
- [2] “Russian Government Decree of 23.10.1993 N 1090 (04.12.2018) «About Traffic Rules»”, *ConsultantPlus* (In Russian).
- [3] Menkin M.I., “Objects model of rules of the road”, *Intelligent systems*, **23**:1 (2019), 39–53 (In Russian).
- [4] Menkin M.I., “Auto-translation of rules of the road to formal graph-theoretic model”, *Intelligent systems*, **23**:3 (2019), 7–31 (In Russian).
- [5] “ABBYY InfoExtractor Demo”.
<http://parsertest.abbyy.com/ru/> (In Russian).
- [6] “Ontology «Linguistic» (ABBYY)”, http://comprenoservice.abbyy.com/help/ru/ontologies_linguistic_linguistic.html (In Russian).