

П. А. Гудков, Е. М. Подмарькова

МОДЕЛЬ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ В ОБЛАСТИ ПРАВОВОЙ ИНФОРМАЦИИ

Аннотация.

Актуальность и цели. Актуальность работы обусловлена необходимостью автоматизации процесса принятия решений по юридическим вопросам в различных областях человеческой деятельности. Цель статьи – продемонстрировать новый подход к организации и структурированию знаний в области правовой информации на основе семантических сетей.

Материалы и методы. Рассматриваются основные методы представления знаний в интеллектуальных системах. В качестве математического аппарата используется теория графов, а также методы Text Mining для извлечения информации из текстовых документов.

Результаты. Показаны преимущества применения семантических сетей как модели представления знаний. Представлен подход к автоматизированному формированию базы знаний по имеющимся текстам правовых документов. Приведен пример практического его применения.

Выводы. Описанная модель открывает широкие перспективы по автоматизации обработки правовой информации. Помимо решения часто встречающихся на практике задач принятия решений по юридическим вопросам, рассмотренный подход может применяться и для такой трудоемкой задачи, как автоматизация проведения юридической экспертизы нормативно-правовых актов.

Ключевые слова: модель представления знаний, семантическая сеть, структурирование, автоматизация, правовая информация.

P. A. Gudkov, E. M. Podmar'kova

LEGAL KNOWLEDGE REPRESENTATION MODEL

Abstract.

Background. The need to automate the decision-making process on legal issues in various fields of human activity determines the importance of this work. The purpose of the article is to demonstrate a new approach to organizing and structuring legal knowledge using semantic networks.

Materials and methods. The basic methods of knowledge representation are considered. Graph theory as a mathematical apparatus, as well as Text Mining methods for extracting information from text documents were used.

Results. The advantages of semantic networks as a knowledge representation model are shown. An approach to the automated formation of a knowledge base based on legal text documents is presented. An example of its practical application is given.

Conclusions. The described model opens up broad prospects for the automation of legal information processing. The considered approach can be used both for solving frequently encountered practical decision-making legal issues, and for the time-consuming task of automating the regulatory legal acts examination.

Keywords: knowledge representation model, semantic network, structuring, automating, legal knowledge.

Введение

Государство в лице своих федеральных органов власти, органов государственной власти субъектов Российской Федерации и органов местного самоуправления осуществляет нормотворческую деятельность по вопросам, отнесенным к их исключительному ведению, и по предметам совместного ведения России и ее территориальных единиц [1].

Все нормотворческие акты хранятся в виде систематизированной совокупности документов законодательства РФ в информационно-справочных системах. В них вся информация о нормативных правовых актах, судебной практике, комментариях и т.д. аккумулирована и позволяет регулировать отношения в правовой сфере деятельности. В качестве примеров существующих информационно-справочных систем можно привести «Консультант Плюс», «Гарант», «Кодекс», «ЮСИС» и др.

Эти справочно-правовые системы представляют собой полнотекстовые базы данных действующего законодательства Российской Федерации со встроенным аппаратом поиска актуальной информации, ежедневно пополняемой и обновляемой. Вся совокупность сведений в них представлена однородными текстовыми данными. Для того чтобы появилась возможность автоматизировать процесс принятия решений по юридическим вопросам в конкретных областях человеческой деятельности, использования одной лишь базы данных с ее жестко заданной структурой недостаточно. Необходим инструмент, который позволит проводить осмысленную обработку имеющейся информации и представлять ее в виде логического вывода по конкретному вопросу, при этом сводя воедино всю совокупность фактов и правил вывода по нескольким правовым документам. Таким решением может служить база знаний. Поэтому возникает необходимость в выборе модели представления знаний в области правовой информации, а также ее практической реализации.

Изучению моделей представления знаний посвящено одно из направлений искусственного интеллекта – инженерия знаний, основной задачей которого является организация имеющихся знаний в формализованное представление базы знаний. Базу знаний можно рассматривать как разновидность базы данных, содержащей структурированную информацию о предметной области. Выбор подходящей модели представления знаний является ключевой проблемой при разработке любых систем.

Проблема адекватного представления знаний обусловлена тем, что в большинстве случаев знания являются продуктом человеческой деятельности. Поэтому основная задача: хранить знания таким образом, чтобы программа могла работать с ними подобно человеку-эксперту, решая тем самым интеллектуальные задачи. Кроме того, работа с естественным языком выдвигает требование аккумулирования всей информации в единую модель представления знаний, что решит проблему поиска знаний в разнородных источниках и существенно облегчит задачу объединения таких источников данных [2, 3].

1. Обзор моделей представления знаний

Все существующие модели представления знаний можно разделить на две группы – логические и эмпирические [4, 5]. Логические модели в данной работе рассматривать не будем в связи с их высокой степенью формализации, что неприменимо к области правовой информации (все законы, распоряжения, указы, постановления формулируются на естественном языке, что предполагает возможность их различного толкования).

Эмпирические модели представления знаний, в свою очередь, делятся на несколько групп, среди которых можно выделить продукционные модели, фреймовые модели и семантические сети. Рассмотрим их более подробно.

1.1. Продукционная модель

В продукционной модели знания представлены в виде правил «ЕСЛИ-ТО». Левая часть правила является условием, которое циклически проверяется интерпретатором, а правая часть – некоторой процедурой, называемой телом правила.

Механизм вывода ищет правила, условия которых соответствуют текущему состоянию базы данных, а затем выполняет их. Правая часть правила обычно отвечает за внесение некоторых изменений в базу данных, добавляя и удаляя элементы. Все правила слабо связаны между собой, поскольку одни из них не могут напрямую вызвать другие правила – все взаимодействие происходит через базу данных [6].

Помимо интерпретатора, в продукционной модели должен быть предусмотрен механизм разрешения конфликтов, который используется для обработки ситуаций, когда более чем одно правило соответствует текущему состоянию базы данных.

Формально продукционная модель может быть представлена в следующем виде [7]:

$$\langle S; L; A \rightarrow B; Q \rangle,$$

где S – описание класса ситуаций; L – условие, при котором правило активируется; $A \rightarrow B$ – ядро продукции; Q – постусловие.

Продукционная модель традиционно используется для представления знаний в экспертных системах и системах поддержки принятия решений в области проектирования и планирования ресурсов. Она привлекает разработчиков своей наглядностью, высокой модульностью, легкостью внесения дополнений и изменений, простотой механизма логического вывода [8, 9].

1.2. Фреймовая модель

Фреймовая модель представления знаний основывается на понятии фрейма – структуре данных, представляющей прототипную ситуацию [10]. Эта структура представляет собой список слотов и во многом похожа на традиционную запись в базе данных. Слоты фрейма могут содержать следующую информацию:

1. Идентификатор фрейма.
2. Отношение этого фрейма к другим фреймам (например, отношение наследования).
3. Описатели требований к фрейму, по которым будет определяться, соответствует ли новый объект данному прототипу.

4. Присоединенные процедуры – отдельные подпрограммы, определяющие, как именно использовать фрейм. Эти процедуры могут быть привязаны как к отдельным слотам, так и к целым фреймам.

5. Информация по умолчанию. Такие значения слотов считаются истинными, если доказательств обратного не найдено.

Таким образом, слоты фрейма определяют его отношения с другими фреймами. А в целом фреймовая модель представляет собой иерархическое описание взаимосвязанного множества фреймов, которое позволяет хорошо систематизировать иерархию объектов реального мира.

К недостаткам модели следует отнести затруднения, возникающие при обмене большими объемами данных между двумя объектами, так как фрейм в классическом виде не предполагает обмена данными [9].

1.3. Семантические сети

Семантическая сеть – это модель представления знаний предметной области в виде графа, вершинами которого являются понятия, а дугами – отношения между ними [7]. Основное достоинство модели – соответствие современным представлениям об организации долговременной памяти у человека.

Из всех моделей представления знаний данная модель является наиболее общей и универсальной, способной хранить как декларативные, так и процедурные знания. Это обеспечивается наличием следующих свойств модели [11]:

1. Полнота. Модель способна представлять любые знания любого типа, независимо от используемой предметной области.

2. Расширяемость. При появлении новой информации, например уточняющей значение некоторого понятия, эта информация может быть легко добавлена к имеющейся семантической сети.

3. Инвариантность при перефразировании. Выражения, имеющие одинаковое значение, должны иметь одинаковое представление в сети независимо от того, как они указаны на уровне предложения.

4. Непрерывность модели. Близкие по значению понятия должны описываться похожими структурами в семантической сети, и небольшие изменения в значении не должны вызывать серьезных изменений их представления. Точно так же понятия, которые имеют резко отличающиеся значения, должны иметь сильно отличающиеся представления.

Предполагается, что система, использующая эту модель, должна содержать в себе следующие компоненты: синтаксический анализатор; пространство узлов, являющееся семантически-сетевым представлением знаний предметной области; интерпретатор, который выполняет операции в пространстве узлов.

Информация, хранящаяся в базе знаний, может использоваться не только по своему прямому назначению, но и для повышения точности работы синтаксического анализатора, что является также серьезным преимуществом данной модели.

Из недостатков модели можно указать на большую вычислительную сложность поиска и вывода на этой модели, которая существенно возрастает с увеличением размера сети.

2. Предлагаемый подход

Исходя из рассмотренных моделей представления знаний, можно сделать вывод о том, что наиболее подходящими из них для представления знаний в сфере обработки правовой информации являются семантические сети.

Авторами предлагается следующий подход для автоматизированного формирования базы знаний по имеющимся текстам правовых документов:

1. Получение перечня исходных текстовых документов, которые описывают конкретную область применения и на основании которых будет строиться семантическая сеть.

2. Применение методов Text Mining для автоматического разбора входных текстовых данных.

3. Автоматизированный контроль выделенных в тексте понятий.

4. Автоматическое или автоматизированное формирование отношений между выделенными понятиями.

В итоге получается граф, вершинами которого являются понятия предметной области.

Рассмотрим пример построения семантической сети для Конституции Российской Федерации. Конституция представляет собой текстовый документ, разделенный на отдельные статьи, являющиеся законченными логическими блоками. Отдельные статьи могут еще делиться на пункты и подпункты. Соответственно методы Text Mining будем применять относительно этих блоков, каждый из которых представляет собой одно или два предложения.

После того как основные понятия выделены, необходим этап их автоматизированного контроля. Это связано с тем, что методы автоматического синтаксического разбора естественного языка не всегда дают на выходе правильный результат. Например, в одной из статей Конституции указывается на внесение предложений одной пятой частью членов Совета Федерации. Слово «пятой» было распознано некорректно – как пята, а не как одна пятая.

Другая причина необходимости автоматизированного контроля найденных в тексте понятий связана с тем, что зачастую понятия образуют не только слова, но и некоторые устойчивые словосочетания:

- Российская Федерация – встречается в тексте более 200 раз,
- Президент Российской Федерации – 110 раз,
- Государственная Дума – 104 раза,
- Федеральный закон – 88 раз,
- Совет Федерации – 72 раза,
- Субъект Российской Федерации – 67 раз,
- Правительство Российской Федерации – 59 раз,
- Конституция Российской Федерации – 53 раза,
- и т.д.

Такие устойчивые словосочетания находятся автоматически, задача пользователя – лишь проконтролировать, что все понятия определены правильно. На текущий момент реализован автоматический поиск словосочетаний из двух, трех и четырех слов.

Нормализация словоформ также выполняется полностью автоматически. Например, дано утверждение: «Право частной собственности охраняется законом». Для него будет выделено три понятия: закон, право, собственность.

На графе семантической сети эти понятия будут представлены отдельными вершинами.

Другой пример: «Президент Российской Федерации обладает неприкосновенностью». Тут будет выделено всего две вершины: «неприкосновенность» и «Президент Российской Федерации».

Заключительный этап предлагаемого подхода – формирование отношений между выделенными понятиями. Данную задачу можно решить несколькими способами. Классический вариант – определение отношений, исходя из выполненного синтаксического анализа предложения. Недостаток подхода – высокая трудоемкость для человека-эксперта в связи с большим количеством неточностей при выполнении автоматического синтаксического разбора предложения.

При большом объеме входных текстовых данных указанный недостаток становится существенным. Поэтому в тех случаях, когда из всего множества отношений требуются лишь отношения зависимости, авторы предлагают использовать следующий подход.

Будем исходить из предположения, что все понятия, найденные в общем блоке текста, являются взаимосвязанными. То есть на всей семантической сети выделяется подграф, вершины которого взаимосвязаны. Сила связи обратно пропорциональна количеству связей в подграфе (которое зависит от количества вершин):

$$f = 2 / (N(N-1)),$$

где N – число вершин в подграфе.

Если какие-то вершины входят одновременно в несколько подграфов, то веса связей суммируются. На рис. 1 представлен фрагмент получившейся семантической сети. Как видно из рис. 1, вершины «закон» и «право» входят в два подграфа, что приводит к суммированию весов связей:

$$f_1 = 2 / (3 \cdot 2) = 0,33 \text{ (Статья 35.1 – 3 вершины в подграфе);}$$

$$f_2 = 2 / (5 \cdot 4) = 0,1 \text{ (Статья 48.1 – 5 вершин в подграфе).}$$

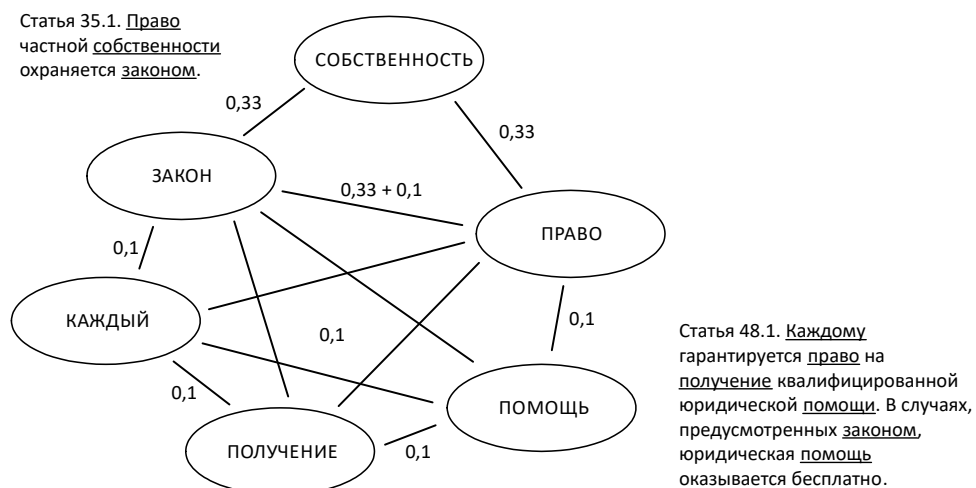
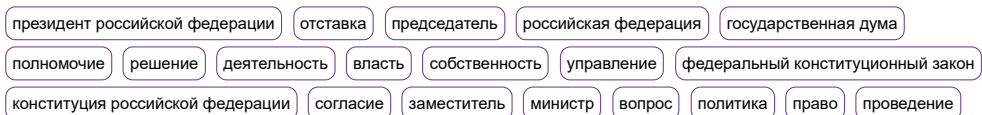


Рис. 1. Фрагмент семантической сети, демонстрирующий суммирование весов отдельных связей

Для проверки того, что предложенный подход применим на практике, на кафедре «Системы автоматизированного проектирования» Пензенского государственного университета была разработана информационно-справочная система, позволяющая выполнять удобную навигацию по отдельным статьям Конституции.

Система представляет собой web-сайт, отдельные страницы которого соответствуют выделенным понятиям предметной области. Страницы соединяются друг с другом гиперссылками, которые соответствуют связям семантической сети (рис. 2).

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ



- Статья 110.1 Исполнительную [власть Российской Федерации](#) осуществляет [Правительство Российской Федерации](#).
- Статья 114.1.г [Правительство Российской Федерации](#) осуществляет [управление](#) федеральной [собственностью](#).
- Статья 114.2 Порядок [деятельности](#) [Правительства Российской Федерации](#) определяется [федеральным конституционным законом](#).
- Статья 116 Перед вновь избранным [Президентом Российской Федерации](#) [Правительство Российской Федерации](#) слагает свои [полномочия](#).
- Статья 117.1 [Правительство Российской Федерации](#) может подать в [отставку](#), которая принимается или отклоняется [Президентом Российской Федерации](#).
- Статья 83.б [Президент Российской Федерации](#) имеет [право](#) председательствовать на [заседаниях](#) [Правительства Российской Федерации](#).
- Статья 83.в [Президент Российской Федерации](#) принимает [решение](#) об [отставке](#) [Правительства Российской Федерации](#).

Рис. 2. Визуализация одной из вершин семантической сети в виде гипертекстового документа

Заключение

Описанная в статье модель открывает широкие перспективы по автоматизации обработки правовой информации. Помимо решения часто встречающихся на практике задач принятия решений по юридическим вопросам, рассмотренный подход может применяться и для такой трудоемкой задачи, как автоматизация проведения юридической экспертизы нормативно-правовых актов. Данный вид экспертизы направлен на выявление различного рода ошибок, на несоответствие Основному Закону РФ (Конституции РФ) и актам большей юридической силы. Поскольку юридическая экспертиза проводится с целью ликвидации информационной недостаточности, неопределенности, с которой сталкиваются субъекты правовых отношений, то она является фактором эффективности правотворчества, правореализации и правоинтерпретации, вспомогательным компонентом при принятии юридически значимых решений [1].

Разработка в дальнейшем полноценной базы знаний предоставит возможности автоматизации труда специалистов юридической сферы, что существенно сократит временные затраты и повысит качество принимаемых правовых решений.

Библиографический список

1. **Пирбудагова, Д. Ш.** Юридическая экспертиза нормативных правовых актов: учеб. пособие / под ред. Д. Ш. Пирбудагова. – 2-е изд., перераб. и доп. – Махачкала : Изд-во ДГУ, 2017. – 199 с.
2. **Верхотурова, Ю. С.** Онтология как модель представления знаний / Ю. С. Верхотурова // Вестник Бурятского государственного университета. – 2012. – № 15. – С. 32–37.
3. **Норvig, П.** Искусственный интеллект: современный подход / П. Норvig, С. Рассел. – Москва : Вильямс, 2007. – 482 с.
4. **Морозова, В. А.** Представление знаний в экспертных системах : учеб. пособие / В. А. Морозова, В. И. Паутов. – Екатеринбург : Изд-во Урал. ун-та, 2017. – 120 с.
5. **Попов, Э. В.** Экспертные системы. Решение неформализованных задач в диалоге с ЭВМ / Э. В. Попов. – Москва : Наука, 1987. – 288 с.
6. **Лорьер, Ж.-Л.** Системы искусственного интеллекта / пер. с фр. под ред. В. Л. Стефанюка / Лорьер, Ж.-Л. – Москва : Мир, 1991. – 568 с.
7. **Гаврилова, Т. А.** Базы знаний интеллектуальных систем : учебник для вузов / Т. А. Гаврилова, В. Ф. Хорошевский. – Санкт-Петербург : Питер, 2000. – 384 с.
8. **Смолин, Д. В.** Введение в искусственный интеллект: конспект лекций : учеб. пособие / Д. В. Смолин. – 2-е изд., перераб. – Москва : Физматлит, 2007. – 264 с.
9. **Толок, А. В.** Обзор методов и моделей представления знаний в САПР / А. В. Толок, М. В. Давидовский // Системы проектирования технологической подготовки производства и управления этапами жизненного цикла промышленного продукта : тр. 14-й Междунар. конф. (CAD/CAM/PDM-2014, Москва). – Москва : Аналитик, 2014. – С. 69–73.
10. **Minsky, M. A.** Framework for Representing Knowledge / M. A. Minsky // The Psychology of Computer Vision / Patrick Henry Winston (ed.). – McGraw-Hill, New York (U.S.A.), 1975.
11. **Feigenbaum, E.** The Handbook of Artificial Intelligence / E. Feigenbaum, Avron Barr. – Addison-Wesley, 1986. – Vol. III.

References

1. Pirbudagova D. Sh. *Yuridicheskaya ekspertiza normativnykh pravovykh aktov: uchebnoe posobie* [Legal expertise of regulatory legal acts: teaching aid]. 2nd ed., rev. and suppl. Makhachkala: Izd-vo DGU, 2017, 199 p. [In Russian]
2. Verkhoturova Yu. S. *Vestnik Buryatskogo gosudarstvennogo universiteta* [Bulletin of Buryat State University]. 2012, no. 15, pp. 32–37. [In Russian]
3. Norvig P., Russell S. *Iskusstvennyy intellekt: sovremennyy podkhod* [Artificial intelligence: a modern approach]. Moscow: Vil'yams, 2007, 482 p. [In Russian]
4. Morozova V. A., Pautov V. I. *Predstavlenie znaniy v ekspertnykh sistemakh: uchebnoe posobie* [Knowledge representation in expert systems: teaching aid]. Ekaterinburg: Izd-vo Ural. un-ta, 2017, 120 p. [In Russian]
5. Popov E. V. *Ekspertnye sistemy. Reshenie neformalizovannykh zadach v dialoge s EVM* [Expert systems. Solving non-formalized tasks in a dialogue with a computer]. Moscow: Nauka, 1987, 288 p. [In Russian]
6. Lor'er Zh.-L. *Sistemy iskusstvennogo intellekta* [Artificial intelligence systems]. Transl. from French., ed by. V. L. Stefanyuk. Moscow: Mir, 1991, 568 p. [In Russian]
7. Gavrilova T. A., Khoroshevskiy V. F. *Bazy znaniy intellektual'nykh sistem: uchebnyk dlya vuzov* [Knowledge bases of intelligent systems: textbook for universities]. Saint-Petersburg: Piter, 2000, 384 p. [In Russian]
8. Smolin D. V. *Vvedenie v iskusstvennyy intellekt: konspekt lektsiy: ucheb. posobie* [Introduction to artificial intelligence: lecture notes: teaching aid]. 2nd ed., rev. Moscow: Fizmatlit, 2007, 264 p. [In Russian]

9. Tolok A. V., Davidovskiy M. V. *Sistemy proektirovaniya tekhnologicheskoy podgotovki proizvodstva i upravleniya etapami zhiznennogo tsikla promyshlennogo produkta: tr. 14-y Mezhdunar. konf. (CAD/CAM/PDM-2014, Moskva)* [Design systems for technological preparation of production and management of stages of the life cycle of an industrial product: proceedings of the 14th International conference (CAD/CAM/PDM-2014, Moscow)]. Moscow: Analitik, 2014, pp. 69–73. [In Russian]
10. Minsky M. A. *The Psychology of Computer Vision*. Patrick Henry Winston (ed.). McGraw-Hill, New York (U.S.A.), 1975.
11. Feigenbaum E., Avron Barr *The Handbook of Artificial Intelligence*. Addison-Wesley, 1986, vol. III.

Гудков Павел Анатольевич

кандидат технических наук, доцент,
кафедра систем автоматизированного
проектирования, Пензенский
государственный университет
(Россия, г. Пенза, Красная, 40)

E-mail: p.a.gudkov@gmail.com

Gudkov Pavel Anatol'evich

Candidate of engineering sciences,
associate professor, sub-department
of computer aided design systems,
Penza State Univeristy (40 Krasnaya
street, Penza, Russia)

Подмарькова Екатерина Михайловна

кандидат технических наук, доцент,
кафедра систем автоматизированного
проектирования, Пензенский
государственный университет
(Россия, г. Пенза, Красная, 40)

E-mail: alpha-and-amega@yandex.ru

Podmar'kova Ekaterina Mikhaylovna

Candidate of engineering sciences,
associate professor, sub-department
of computer aided design systems,
Penza State Univeristy (40 Krasnaya
street, Penza, Russia)

Образец цитирования:

Гудков, П. А. Модель представления знаний в области правовой информации / П. А. Гудков, Е. М. Подмарькова // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2020. – № 3 (55). – С. 17–25. – DOI 10.21685/2072-3059-2020-3-2.