

УДК 340

### ПОСТРОЕНИЕ ЦИФРОВОЙ МОДЕЛИ ПРАВОВОГО ПРОСТРАНСТВА И ЕЕ ИСПОЛЬЗОВАНИЕ ДЛЯ РЕШЕНИЯ ЮРИДИЧЕСКИХ ЗАДАЧ

© 2019

**В. В. Крылов<sup>1</sup>, И. А. Филипова<sup>2</sup>, Г. М. Жигалов<sup>1</sup>**

<sup>1</sup> Национальный исследовательский университет «Высшая школа экономики»

<sup>2</sup> Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского

Правовые нормы, определяющие пространство права, являются языковыми конструкциями, построенными в соответствии с правилами естественного языка. Однако для их правильного формулирования и применения требуется знание правил юридической техники, то есть специальное обучение, которое, в отличие от иных интеллектуальных областей деятельности человека, пока не использует язык математики. В настоящем исследовании представлены результаты построения эмпирической математической модели правового пространства на примере норм российского уголовного и гражданского права, а также анализ возможностей по интерпретации и созданию приложений, в которых такая модель может использоваться. Авторами демонстрируется, что математическое «правовое пространство» является обособленным кластером естественного языка, объективные понятия близости в таком пространстве порождают понятия связанности юридических утверждений, давая возможность прогнозировать правовую квалификацию фактов, описание которых сформулировано на естественном языке, с высокой степенью достоверности. В целях иллюстрации возможностей использования представленной модели разработан классификатор оценки «успешности» гражданского дела и построен прототип чат-бота – юриста-помощника для интерпретации произвольного высказывания в терминах, свойственных юридической технике, и квалификации деяний по нормам права.

*Ключевые слова:* цифровая модель правового пространства, искусственный интеллект, правоприменение, машинное обучение.



**В. В. Крылов**

*Профессор кафедры прикладной математики и информатики Национального исследовательского университета «Высшая школа экономики», доктор технических наук, профессор*



**И. А. Филипова**

*Доцент кафедры гражданского права и процесса юридического факультета Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского, кандидат юридических наук, доцент*



**Г. М. Жигалов**

*Магистрант Национального исследовательского университета «Высшая школа экономики»*

Правовые нормы определяют ограничения в отношении действий членов общества, тем самым способствуя формированию устойчивости общественного устройства и стабильности существования его биологических носителей. Ускоряющаяся динамика развития современного общества требует быстрой и качественной работы правовых механизмов. Сложность и важность конструирования новых правовых норм (правотворчества) и правильной интерпретации имеющихся норм (правоприменения) заставляет искать новые пути решения имеющихся в этой сфере проблем<sup>1</sup>. Подобно другим областям интеллектуальной деятельности прогресс в решении имеющихся проблем многие современные исследователи связывают с внедрением технологий искусственного интеллекта (далее – ИИ). Однако в большинстве областей, где ИИ уже успешно используется, предстоящее внедрение цифровых технологий предварительно исследовалось научными методами, опирающимися на математические модели<sup>2</sup>.

При проведении исследований в области права математические (цифровые) модели используются в очень ограниченном объеме или не используются совсем. Отчасти это объясняется консерватизмом юридической науки и практики и, к сожалению, сложностью математических моделей для понимания многими представителями юриспруденции. А это, в свою очередь, существенно затрудняет разработки в области ИИ для применения в правовой сфере и, как правило, сводит их к упрощению работы с документами или к созданию вопросно-ответных систем общего вида.

Означает ли это, что системы ИИ пока не стоит развивать в правовой сфере? Ответ будет отрицательным, хотя бы потому, что в настоящее время технологии ИИ начали использовать в управлении, в том числе в публичном управлении – сфере, где правовое регулирование играет критически важную роль. Складывающаяся тенденция к устойчивому развитию ИИ в области управления усиливает проблему ограниченности применения системы существующих правовых норм к управленческим отношениям, нацеливая на поиск новых решений. Следующий этап – внедрение ИИ в правоприменение. То есть внедрение технологий ИИ на сегодня становится мегатрендом, охватывающим промышленность («Индустрия 4.0»), образование, здравоохранение, управление («электронное правительство») и иные сферы жизни общества. Право как социальный регулятор не сможет «выключиться» из общего потока изменений, оста-

ваясь при этом социальным регулятором. Правовое регулирование не может игнорировать общий темп технологического прогресса.

Итак, начнем с относительно простого: используя метод математического моделирования, попробуем построить модель правового пространства на примере норм российского уголовного и гражданского права. Как может выглядеть математическая модель права, представленного в реальной жизни множеством документов? Тривиальная модель в виде дискретного множества представляется бесполезной, нужно найти такое сужение модели, которое приобрело бы практический смысл. Применение формальных таксономий и иерархий, обычно используемых в классификации отдельных документов, не приближает к решению задачи. Знание того, к какой части Гражданского кодекса РФ принадлежит каждая статья, к какому разделу права относится принятое судебное решение, все равно требует плохо формализуемых знаний юриста, чтобы правильно интерпретировать тот или иной набор фактов, описывающих реальную жизнь и имеющих юридическое значение.

Рассуждения о праве как абстрактном математическом объекте представлялись авторам исследования неконструктивными, и было принято решение использовать так называемый *data driven* подход к построению модели, основанный на реальных данных<sup>3</sup>. В процессе были использованы полные тексты Уголовного кодекса РФ и Гражданского кодекса РФ (каждая статья учитывалась как самостоятельная единица, итого: 496 и 1151 документ). Однако реальное правовое пространство опирается не только на нормативные правовые акты, но и на практику, поэтому материалом для построения модели послужили и тексты судебных решений по уголовным делам (17 тыс. документов) и по гражданским делам (10 тыс. документов). С самого начала было очевидно, что хотя все юридические документы пишутся на естественном языке, используемые в них слова опираются на внешние по отношению к праву понятия. Поэтому при построении модели был использован набор документов, которыми принято определять полный русский язык, в данном случае – открытый корпус русского языка из 4000 текстов (*OpenCorpora*). Подобная база текстов на русском языке может служить достаточной эмпирической основой для построения математической модели.

В качестве математической модели собранного таким образом корпуса текстов было принято решение применить машинное обучение через широ-

ко использующиеся в настоящее время эмбединги (embeddings) – представления в виде числовых векторов высокой размерности. В числовых векторных пространствах определены множества привычных для математиков понятий: расстояния, линейные и нелинейные преобразования, известно, как визуализировать такие пространства. Итак, постановка задачи построения модели правового пространства теперь формулируется следующим образом: построить представление данного текстового корпуса в многомерном векторном пространстве с сохранением базовых семантических отношений и инвариантов. Обычно это иллюстрируется примером: слова «король» и «королева» должны иметь такие векторы представления, что их разность совпадает с разностью векторов для слов «мужчина» и «женщина».

При использовании векторного представления (эмбединга) правового пространства серьезной проблемой стала необходимость решить, какие языковые единицы следует отображать в вектор и как выбрать размерность векторного пространства эмбедингов. При описании были использованы тексты, очищенные от именованных сущностей (имена, даты, географические названия) и постоянно сопровождающих тексты стоп-слов. Из судебных решений для эмбединга выбиралась только описательная часть. После очистки каждый текст преобразовывался в список токенов, которые далее приводились к базовой словоформе: устранялись падежи, склонения и т.п. На последнем этапе подготовки токены объединялись в так называемые коллокации для нормализации полученных списков.

Главный этап отображения списка токенов в вектор был осуществлен с помощью алгоритма машинного обучения, называемого doc2vec. Этот алгоритм основан на анализе вероятностей соседних слов с преобразуемым словом. Визуализация полученно-

го векторного пространства позволила увидеть ряд особенностей «правового пространства» и его составляющих, относительно обособленность от классического русского языка. На рисунке 1 приведено взаимоотношение нормативных правовых актов, а точнее – статей Уголовного кодекса РФ, и актов судебной практики.

Как видно на рисунке, статьи Уголовного кодекса РФ в целом формируют ядро принимаемых судебных решений, но среди пула судебных документов видны два фокуса, один из которых согласуется с Уголовным кодексом РФ, другой, хотя и малочисленнее, образует некоторое обособленное подмножество.

Аналогичная картина просматривается и при рассмотрении норм гражданского права и соответствующей судебной практики. Гражданский кодекс РФ в целом совпадает по нахождению с «ядром» судебных решений в этой области, однако и здесь есть обособленный кластер решений (рис. 2).

Определенный интерес вызывает наблюдение за тем, как язык юристов отличается от классического русского языка, почему юридические документы так трудно прочитать и понять. На следующем рисунке можно увидеть, насколько сильно классический русский язык отличается от языка уголовного права (рис. 3).

Наконец, если визуализировать все рассматриваемые компоненты права и русский язык классиков, то картина будет выглядеть следующим образом (рис. 4).

Язык гражданского права не менее отличается от классического русского языка. На рисунке 4 это хорошо видно.

Построенное векторное пространство наделено естественной мерой близости между любой парой точек – евклидовым или любым другим расстоянием, поэтому строгие количественные соотноше-

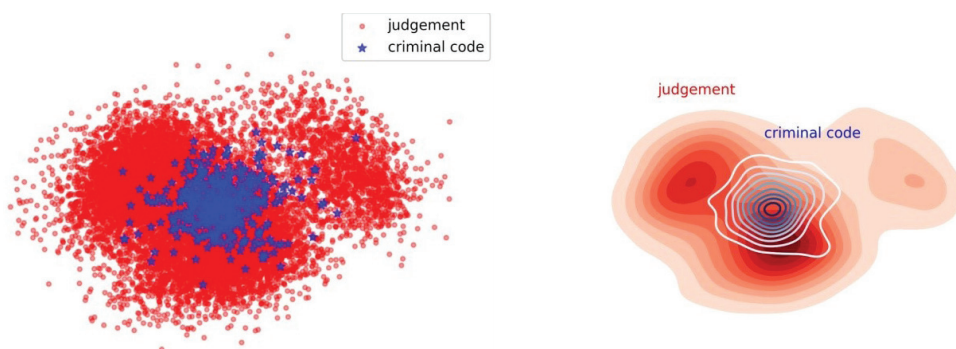


Рис. 1

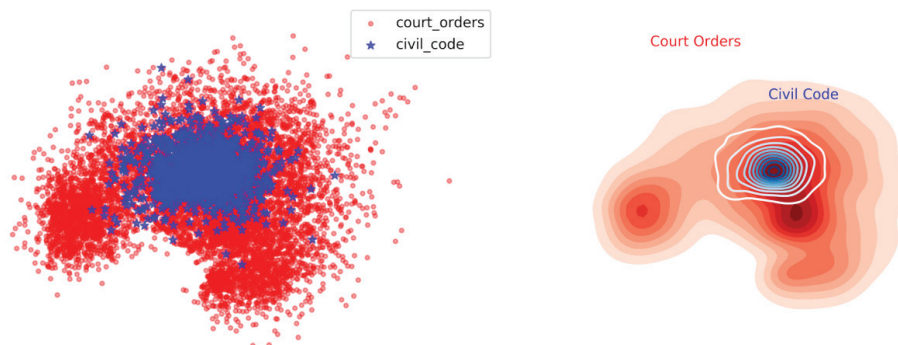


Рис. 2

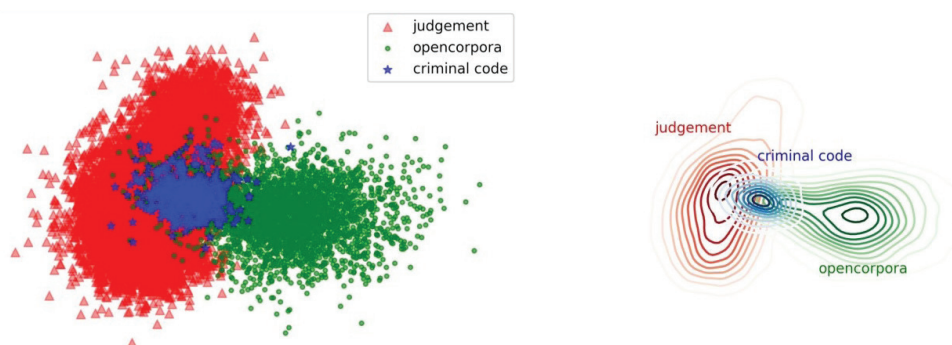


Рис. 3

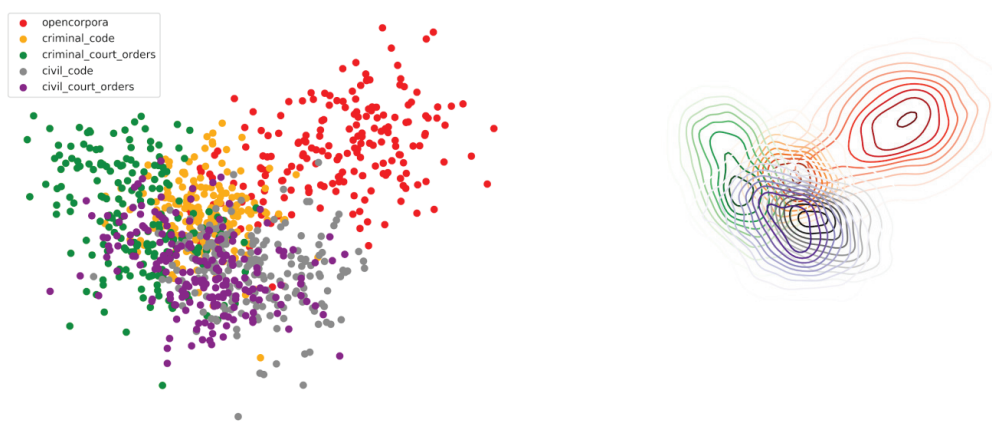


Рис. 4. Взаимное расположение всех рассматриваемых компонентов: Гражданский кодекс РФ, Уголовный кодекс РФ, акты судебной практики по гражданским и уголовным делам и общий русский язык

ния можно установить между любыми документами «правового пространства», которое в данном случае было построено. Близость, похожесть различных документов можно проиллюстрировать, создав модель в виде ненаправленного графа, вершинам которого соответствуют отдельные правовые нормы (например, норма, зафиксированная в части 1 статьи 264 Уголовного кодекса РФ). Поскольку расстояния между векторами могут быть любым неотрицательным числом, то имеет смысл рассмо-

треть так называемые пороговые графы – те, для которых ребро между каждой парой вершин считается существующим, если расстояние между этими вершинами не превышает заданного числа (порога). Тогда для каждого порога мы получаем граф тем более связный, чем большее значение порога выбрано. Значение порога можно трактовать как величину, характеризующую степень важности связей. Чем порог меньше, тем только более существенные связи выделяются.

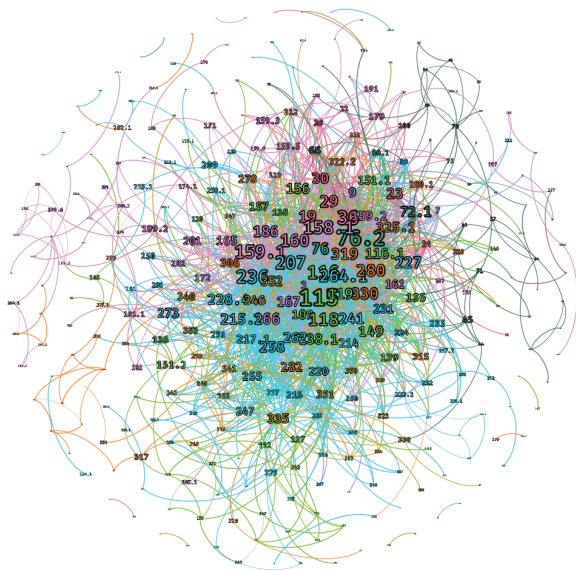


Рис. 5. Порог 0.7: множество ребер, здесь еще трудно увидеть структуру кодекса

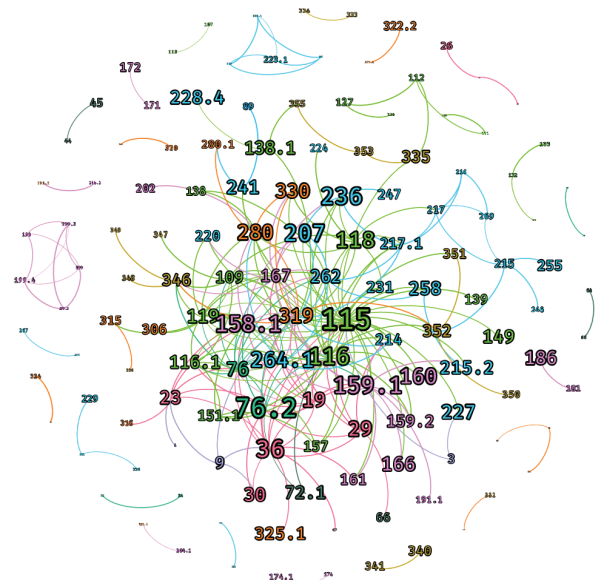


Рис. 6. Порог 0.3: структура уже прослеживается, видно, что некоторые статьи группируются и не связаны с общим ядром

Посмотрим на пороговые графы Уголовного кодекса РФ для трех значений порога: 0.7, 0.3, 0.2 (рис. 5, 6, 7).

Величина порога, которую мы выбирали, имеет чисто инструментальный характер. При выборе других мер расстояния они будут изменяться. Однако распад корпуса текстов Уголовного кодекса РФ на отдельные кластеры при снижении порога является его объективным свойством, и, как видно из анализа полученных графов, существуют достаточно изолированные группы связанных статей, которые сильно похожи внутри этой группы и непохожи на другие. Это говорит о том, что формулировки отдельных групп статей используют отличные от других конструкции русского языка, что может приводить к дополнительным трудностям при изучении и толковании.

На всех рисунках размер шрифта для указания номера статьи пропорционален числу лингвистически похожих других статей Уголовного кодекса РФ на данную. Можно заметить, что наибольшее число похожих на другие имеют два элемента – статьи 115 и 116 Уголовного кодекса РФ. На третьем месте по связанности оказалась статья 76.2 Уголовного кодекса РФ. Объяснение этого факта само по себе будет интересным, но выходит за рамки данного исследования.

Итак, цифровая модель правового пространства построена. Какие задачи можно решать на

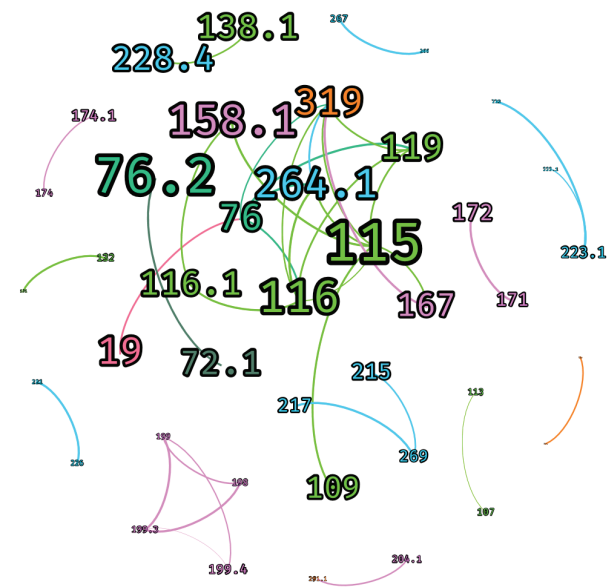


Рис. 7. Порог 0.2: хорошо выделяются отдельные сильно связанные группы разделов Уголовного кодекса РФ

основе данной модели? Одной из задач, которые могут быть решены таким образом, является прогнозирование того, по какой статье Уголовного кодекса РФ будет вынесен обвинительный приговор, если известен текст проекта судебного решения (его описательной части). В качестве эксперимента был написан робот (интернет-бот), который использовал бы описанную выше модель право-

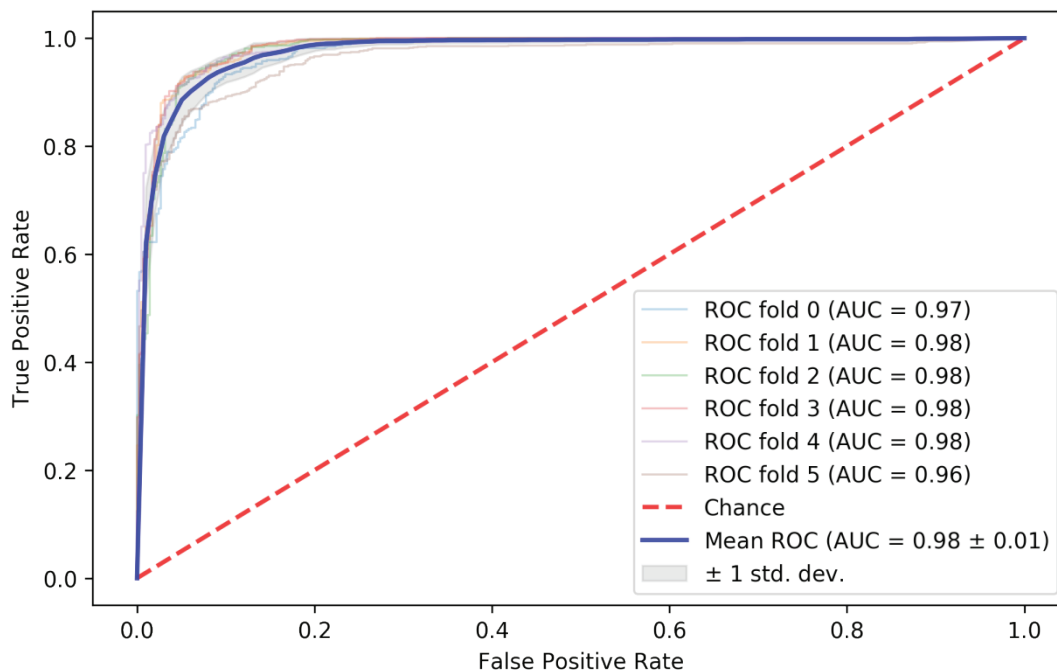


Рис. 8. Характеристика правильности прогнозирования судебного решения по гражданскому иску: 98%

вого пространства. Обучение выполнялось следующим образом: было выделено, по каким статьям выносились обвинительные приговоры. Далее, часть судебных решений была использована для обучения бота: ему предъявлялась описательная часть решения и статья обвинения. Обучение было выполнено на 10 тыс. примеров судебных решений, после чего обученный бот был протестирован на 7 тыс. уже вынесенных решений судов первой инстанции: ему сообщалась только описательная часть, а его задачей было прогнозировать статью обвинения. Для построения такого классификатора на 496 классов (по числу статей Уголовного кодекса РФ) использовался простой алгоритм SVM RBF. Тест показал неплохой результат для такого количества классов: F1-оценка оказалась равной 0.62. Результат явно может быть улучшен, поскольку алгоритм машинного обучения был взят по принципу известности, без предварительного анализа и оптимизации.

Лучшие результаты бот показал для бинарной классификации гражданских дел, когда велось его обучение прогнозированию по описательной части проекта судебного решения: будет ли удовлетворен иск. Здесь точность прогноза оказалась равна 98%. На рисунке 8 приведен график кривой распознавания ROC для проведенных тестов.

Впоследствии было решено представить разработанный бот в открытом доступе. Был написан чат-бот (Telegram@legal\_space\_bot, доступный в настоящее время), способный находить наиболее близкие по смыслу статьи Уголовного кодекса РФ и Гражданского кодекса РФ, под действие которых подпадает описанная пользователем ситуация. В качестве примера можно привести чат (см. рис. 9).

Как видим, данная математическая модель в виде многомерного числового векторного пространства позволяет создавать роботов (ботов) для анализа юридических текстов в сфере уголовного и гражданского права, прогнозирования судебных решений и проведения предварительных юридических консультаций. Аналогичные модели могут быть выстроены и для других отраслей права – трудового, административного и т.д. Это значительно облегчит доступ к праву для неюристов, а также освободит практикующих юристов от больших объемов рутинной работы, высвободив время на решение иных задач. Кроме того, это, с одной стороны, позволяет внедрить новые технологии в правовую сферу, а с другой – соответствует требованиям правового государства, проникновению «духа» права в массовое сознание и повышению уровня правовой культуры общества.

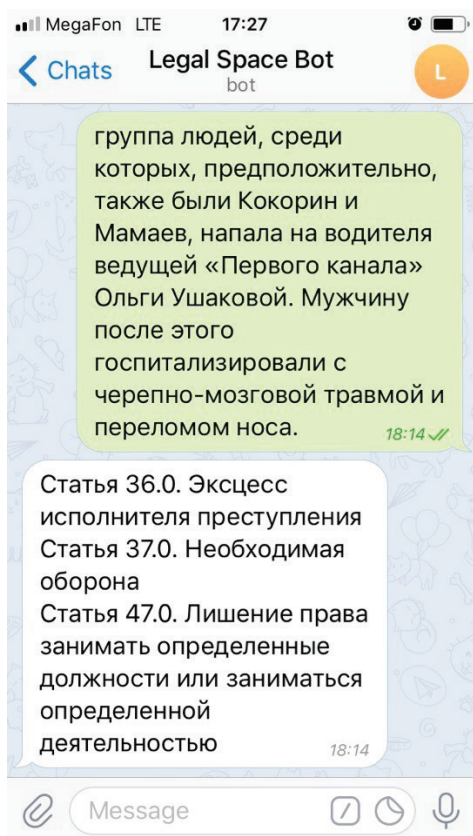


Рис. 9. Скриншот работы чат-бота для поиска подходящих к описанию фактов статей Уголовного кодекса РФ

В заключение следует отметить, что исследование продолжается, авторами планируется изучение новых подходов к построению числовых пространств для юридических документов и расширение используемой базы судебных решений, других материалов правового характера для обучения роботов, ведь цифровизация права подобно цифровизации экономики открывает новые возможности по объективизации проблем в процессе ведения правовых исследований, в правоприменении и правотворчестве, способствуя совершенствованию этих сфер человеческой деятельности. ■

#### *Библиографический список*

1. Талапина Е.В. Право и цифровизация: новые вызовы и перспективы // Журнал российского права. 2018. № 2. С. 5–17.
2. Flach P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data. New York: Cambridge University Press, 2012. 409 p.

<sup>1</sup> См., например: Талапина Е.В. Право и цифровизация: новые вызовы и перспективы // Журнал российского права. 2018. № 2. С. 5–17.

<sup>2</sup> Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-011-00320.

<sup>3</sup> Flach P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data. New York: Cambridge University Press, 2012. 409 p.

## **CONSTRUCTION OF A DIGITAL MODEL OF THE LEGAL SPACE AND USING IT TO SOLVE LEGAL TASKS**

***V. V. Krylov***

Professor of the Department of Applied Mathematics and Computer Science  
of the National Research University «Higher School of Economics» (Nizhny Novgorod), Doctor of Sciences (Engineering)

***I. A. Filipova***

Associate Professor of the Department of Civil Law and Procedure of the Lobachevsky State University of Nizhny Novgorod,  
Candidate of Sciences (Law)

***G. M. Zhigalov***

Master's student of the National Research University «Higher School of Economics»

Legal norms defining the space of law are language constructs built in accordance with the rules of a natural language. Knowledge of the rules of legal technique is required for the proper formulation and application of legal norms. Thus, special training is required, which, unlike many other intellectual fields of human activity, does not yet use the language of mathematics.

This study presents the results of building an empirical mathematical model of the legal space on the example of the norms of Russian criminal and civil law, as well as an analysis of the possibilities for interpreting and creating applications in which such a model can be used. The authors of the study indicate that the mathematical legal space is a separate cluster of natural language, and the objective concepts of proximity in such a space give rise to the concept of connectedness of legal statements. This makes it possible with a high degree of accuracy to predict the results of the legal qualification of the facts, the description of which is formulated in natural language. A classifier assessing the success of a civil case is designed to illustrate the possibilities of using the presented model. In addition, a prototype chatbot (legal assistant) was built to interpret arbitrary statements in legal terms and to qualify acts according to the norms of law.

*Keywords:* digital model of legal space, artificial intelligence, law enforcement, machine learning.