

## ТЕХНОЛОГИЯ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА (NLP) В ЗАКОНОДАТЕЛЬНОМ ПРОЦЕССЕ

**С. С. Зенин, Д. Л. Кутейников**

*Университет имени О. Е. Кутафина (МГЮА), г. Москва,*

**И. М. Япрынецев**

*Конституционный Суд Российской Федерации, г. Москва,*

**О. А. Ижаев**

*Московская городская избирательная комиссия, г. Москва*

В статье обозначена взаимосвязь внедрения технологии машинного обучения в законодательный процесс с необходимостью обработки естественного языка (Natural Language Processing, NLP), которая выступает необходимым элементом в этом процессе.

Формирование баз данных, охватывающих существующий нормативный массив, связано с необходимостью его перевода в машиночитаемый вид, понятный алгоритму. Использование NLP позволяет осуществить такое преобразование.

Авторами описаны примеры использования NLP в правовой сфере. С учетом результатов такого использования в статье оцениваются потенциальные возможности внедрения NLP в рамках законодательной деятельности. При этом отмечается специфика юридических текстов, что предопределяет необходимость дополнительных требований к процессу их обработки.

В заключении авторы отмечают некоторые риски, с которыми сопряжено внедрение машинного обучения в законодательный процесс, но которые могут быть минимизированы при использовании NLP.

**Ключевые слова:** *законодательный процесс, автоматизация, алгоритм, машинное обучение, обработка естественного языка, семантическое значение текста.*

Стремление<sup>1</sup> выработать критерии объективности в законотворческой деятельности, обеспечить механизмы, нивелирующие предвзятость и субъективизм в рамках разработки и принятия правотворческих решений связано с необходимостью устранения последствия произвольных решений [3]. Современные цифровые технологии являются одним из возможных вариантов минимизации негативных проявлений «человеческого фактора» в правотворчестве.

Объективная необходимость перевода законодательных процедур на цифровую платформу продиктована не только активно развивающимися технологиями, которые все в большем объеме используются практически во всех сферах деятельности человека, общества, государства, но и подтвержденными на практике позитивными результатами их использования.

Внедрение цифровых технологий в законодательный процесс ведет не только к оптимизации его процедурной составляющей [7], но и существенно расширяет субъектный состав правотворчества, что ведет к увеличению его эффективности и результативности, а также демократизации всей законодательной деятельности. В частности анализ информации и моделирование процессов с использованием искусственного интеллекта позволяет, помимо прочего, выявить такие сферы общественных отношений, регулятивное воздействие на которые не предполагалось при разработке нормативного правового акта, но в действительности может иметь место, а также определить потенциальное воздействие на иные аспекты деятельности человека, общества и государства [9]. К подобного рода технологиям, очевидно, относится и технология машинного обучения.

<sup>1</sup> Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-29-16214.

*Машинное обучение и законодательный процесс.*

В самом общем виде машинное обучение представляет собой категорию алгоритмов, позволяющих на основе входных данных формировать наиболее точные прогнозы относительного итогового результата [2]. Такие алгоритмы могут использоваться в законодательной деятельности в силу их аналитических и предсказательных возможностей, использование которых способно существенно трансформировать процесс разработки и принятия нормативных правовых актов.

Использование технологий машинного обучения в правотворчестве позволяет не только осуществлять «классификацию» текстов по их содержанию, а потому определять их тематическую и регулятивную направленность, но и создавать такого рода алгоритмы, которые самостоятельно (без участия человека) будут способны на основе входных данных функционировать и прогнозировать значение итогового выходного результата [11]. В этой связи для достижения корректных результатов использования подобных технологий важными являются адекватность, применимость и правильность входных данных [1], на основании которых алгоритмы и будут обучаться.

Обучение алгоритмов предполагает необходимость формирования баз данных, охватывающих действующие нормативные правовые акты и правоприменительную практику. Это в свою очередь предопределяет необходимость формализации текста, его перевода из естественного языка в понятный для алгоритма язык. Проблема трансформации текста в машиночитаемый вид может быть разрешена в рамках обработки естественного языка (Natural Language Processing, NLP).

*NLP в правовой сфере: практика использования и потенциал внедрения в правотворчество.*

Распространение NLP в юридической сфере и существующие результаты использования обуславливают возможность их внедрения и в рамках правотворческого процесса. К направлениям такого использования могут относиться, в частности, выявление семантического значения текста, установление причинно-следственных связей между различными частями нормативных актов, установление их регулятивной направленности.

Сформированные подходы к подготовке и

обработке текстов позволяют выделить необходимые этапы, осуществление которых связано с переходом к цифровому законодательному процессу, например, формирование баз данных, содержащих образцы (маркеры), необходимые для анализа текста.

Сфера применения NLP на сегодняшний день начинается от распознавания речи и перевода текстов до предиктивного ввода текста и выстраивания коммуникации между машиной и человеком. Сферы же использования NLP в праве, таким образом, следующие: поиск информации, релевантной по заданным критериям, анализ контрактных обязательств, подготовка юридических документов, формулирование правовых заключений [4].

Эффективность и действенность NLP в работе с юридическими текстами подтверждается на практике. Например, Ross Intelligence (<https://rossintelligence.com/>) и v Lex (<https://vlex.com/>) в рамках деятельности предлагают интерфейс, способный воспринимать информацию на естественном языке так, будто человек ведет общение с юристом, после чего выдавать релевантные по обозначенным проблемам данные.

NLP связано с применением различных подходов, которые в итоге приводят текст, представленный на естественном языке, к такому виду, который позволяет алгоритму его обрабатывать и извлекать из него необходимые данные. Речь в данном случае идет о токенизации (по предложениям и словам), лемматизации и стемминге, выявлении стоп-слов т.д. При этом применение этих подходов к юридическому языку связано с рядом трудностей. Специфика юридических текстов предопределяет, например, наличие следующих сложностей:

– использование аббревиатур и ссылок, что усложняет токенизацию текста;

– извлечение информации из юридических текстов затруднено в силу их синтаксической сложности и насыщенности;

– несмотря на достаточно ограниченный лексикон юридического языка, смысловое значение может существенно различаться, что снижает точность семантического анализа текста [15, 10].

В этой связи процесс имплементации и использования NLP в правотворческой деятельности предопределяет необходимость дополнительных действий, гарантирующих результативность использования данной техно-

логии и способствующих снижению указанных рисков. Одним из возможных вариантов минимизации обозначенных сложностей выступает предварительное создание специальных баз данных, содержащих в себе маркеры, которые алгоритм будет использовать при синтаксическом анализе (парсинге).

Практический результат такого маркирования имеет свое подтверждение. В частности в центре права имени Лейбница в университете Амстердама была сформирована специальная база образцов, используемых при парсинге при поиске ссылок внутри документа и ссылок на другие документы, а также при классификации (определении семантической принадлежности) текстов. Точность такого поиска достигает 90–95 % [14, 6].

Следующим направлением, связанным с использованием NLP в отношении юридических текстов, является возможность сканирования, кластеризации и классификации данных [16], позволяющих обеспечить анализ поступающих проектов нормативных правовых актов с тем, чтобы установить их тематическую принадлежность, определить необходимость регулирования той или иной сферы общественных отношений, исключить противоречие и дублирование с существующими нормами.

Практический опыт использования алгоритмов в подобных вопросах представлен, в частности, в сфере так называемого contract review (анализа положения договоров). При этом алгоритм работает либо на основе ранее заданных типичных условий договоров, среди которых пользователь выбирает подходящие ему, а также может добавить недостающие условия, что в свою очередь обеспечивает дополнительные возможности для обучения алгоритма. По такому алгоритму работает Kira (<https://kirasystems.com/>).

В сфере правотворческого процесса такой подход применим, если говорить об указании типичных характеристик нормы (отраслевой принадлежности, ее цели – установление нового регулирования либо изменение существующего, именно в ней содержатся новые обязанности, запреты или права, круг субъектов, на которых она распространяется), что позволит алгоритму проанализировать данные на предмет установления взаимосвязанных положений в уже действующих актах и определить возможные риски введения нового регулирования – противоречия, дублирование и т.д.

Кроме того, для сферы правотворчества интерес представляют алгоритмы, которые позволяют при анализе текстовых данных договоров сравнивать их с заранее обозначенными политиками компании. Такой алгоритм предлагается компанией Law Geex (<https://www.lawgeex.com/>). С точки зрения имплементации подобных решений в правотворческую деятельность это позволяет проанализировать данные на предмет их корреляции обозначенным целям законопроекта и тому, какие результаты предполагает его введение. Оценка регулятивного воздействия является одной из важнейших стадий законодательного процесса.

Выявление семантического значения данных для законодательного процесса является крайне важным. Это позволяет, во-первых, установить конкретный структурный элемент нормы, представленной в новых данных, то есть определить, связано ли это с правами, обязанностями, запретами или наказаниями для тех или иных субъектов, что в свою очередь предопределяет возможность формирования причинно-следственных связей как внутри отдельно взятой нормы, нормативного акта, так и между различными текстами [12]. Во-вторых, классификация данных по семантике позволяет установить саму направленность нормы – является она регулятивной (то есть вводит новое правило) или направлена на изменение действующего регулирования (связана с корректировкой или исключением существующих норм) [5].

Для форматирования существующего опыта в машиночитаемый вид, чтобы алгоритм мог использовать его в качестве основы для анализа поступающей новой информации, важен также метод кластеризации, который предполагает выделение ключевых функций (маркеров) уже принятых нормативных правовых актов, что позволяет распределить весь объем данных на множества, например, в зависимости от объекта регулирования или даты принятия акта и т.д. Конечным результатом такого метода является формирование алгоритма, позволяющего определить ключевые функции для любого нового набора данных (вносимого проекта нормативного правового акта) и соотнести его с существующим регулированием в этой сфере [8].

Применение на практике возможностей NLP, связанных с решением проблем формализации юридических текстов, возможно для отдельно взятых сфер правового регулирова-

ния, подвергающихся большей алгоритмизации. В частности, в литературе описывается опыт применения специального алгоритма, позволяющего Голландской налоговой и таможенной службе обрабатывать и анализировать новое законодательство в целях его соотнесения с ее задачами. При этом алгоритм основывается на объектно-ориентированном парсинге и анализе законодательства (Object-oriented Parsing and Analysis of Legislation), который предполагает необходимость поиска понятий в тексте на естественном языке, определяющих сферу его применения [10].

*Заключение.*

Использование технологий машинного обучения в законодательном процессе способно существенно повысить его эффективность как за счет снижения организационных и процедурных издержек, так и путем существенного увеличения привлекаемых к этому процессу субъектов. При этом следует исходить и из того факта, что применение машинного обучения сопряжено с определенными недостатками этой технологии. В большинстве своем машинное обучение предполагает наличие заранее обработанных и классифицированных данных, используемых алгоритмами в качестве тренировочных, обучающих. Такие данные выступают основой для анализа и поиска закономерностей. Недостаточность же таких данных, как результат, может привести к искажению конечных результатов и снижению их точности [17]. Соответственно необходимой предпосылкой использования машинного обучения в рамках разработки и принятия нормативных правовых актов является форматирование существующих актов в машиночитаемый вид.

Именно с решением этой задачи и связана обработка естественного языка, по результатам которой алгоритм способен анализировать поступающие данные. Это позволяет охватить целый спектр вопросов, начиная от определения цели правового акта, уяснения его структуры и базовых элементов, до определения внутренних связей между его положениями и связей с иными нормативными актами.

Подходы, которые используются в рамках NLP к анализу иных текстов, применимы и к сфере законодательной деятельности. Особенности же юридических текстов определяют необходимость внедрения дополнительных требований к этой деятельности, не-

которые из которых представлены в настоящей статье.

*Литература*

1. Alpaydin E. Introduction to Machine Learning. 3rd ed. / Ethem Alpaydin. – Cambridge: The MIT Press, 2014.
2. Anthony J. Casey and Anthony Niblett. A Framework for the New Personalization of Law // University of Chicago Law Review, 2019, Vol. 86.
3. Anthony J. Casey and Anthony Niblett. Self-driving Laws // The University of Toronto Law Journal, 2016, Vol. 66, no. 4, P. 437.
4. Dale R. Law and Word Order: NLP in Legal Tech // Natural Language Engineering, 2019, no. 25 (1).
5. De Maat E. and Winkels R. (2008). Automatic classification of sentences in Dutch laws // Legal Knowledge and Information Systems. Jurix 2008: The 21st Annual Conference, volume 189 of Frontiers in Artificial Intelligence and Applications, pp. 211–212.
6. De Maat E., Winkels R., van Engers T. Automated Detection of Reference Structures in Law // Legal Knowledge and Information Systems – JURIX 2006: The Nineteenth Annual Conference on Legal Knowledge and Information Systems, Paris, France, 7–9 December 2006.
7. Metsker O. G., Trofimov E., Petrov M., Butakov N. A. Russian Court Decisions Data Analysis Using Distributed Computing and Machine Learning to Improve Lawmaking and Law Enforcement. Procedia Computer Science, 2019, Vol. 156, P. 265.
8. Noh H., Jo Y., Lee S. Keyword selection and processing strategy for applying text mining to patent analysis // Expert Systems with Applications, 2015, no. 42(9), pp.
9. Oleg Metsker, Egor Trofomov. Text and Data Mining Techniques in Judgment Open Data Analysis for Administrative Practice Control. Electron Gov Open Soc Challenges Eurasia 2019.
10. R. Van Gog and T. M. Van Engers. Modeling legislation using natural language processing // 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat.No.01CH37236), Tucson, AZ, USA, 2001.
11. Silver, D. Mastering the Game of Go with Deep Neural Networks and Tree Search / David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den

Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, NalKalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, KorayKavukcuoglu, ThoreGraepel, Demis Hassabis // *Nature*, 2016, Vol. 529, pp. 484–503.

12. Soria C., Bartolini R., Lenci A., Montemagni S., and Pirrelli V. Automatic extraction of semantics in law documents. *Proceedings of the V Legislative XML Workshop. 2007*, pp. 253–266.

13. Spivak A.I., Razumovskii A.V., Nasonov D.A., Bukhanovsky A.V., Redice A. Storage tier-aware replicative data reorganization with prioritization for efficient workload processing // *Future Generation Computer Systems, IET. 2018. Vol. 79. Part. 2.*

14. Teaching The Computer To Read Legal Text. URL: <https://blog.law.cornell.edu/voxpop/tag/legal-natural-language-processing>.

15. Using NLP and ML to Analyze Legislative Burdens Upon Businesses (Режимдоступа – <https://medium.com/@ODSC/using-nlp-and-ml-to-analyze-legislative-burdens-upon-businesses-e5cc106b85b0>).

16. Wafapoor V. Natural Language Processing: An Essential Tool for Attorneys (Режимдоступа: <https://www.lawpracticetoday.org/article/natural-language-processing-tool-attorneys/>).

17. Saarikoski J. et al. On the Influence of Training Data Quality on Text Document Classification Using Machine Learning Methods // *Int. J. Knowl. Eng. Data Min. 2015. № 2 (3). pp. 143–169.*

**Зенин Сергей Сергеевич** – кандидат юридических наук, директор, НИИ, Университет имени О. Е. Кутафина (МГЮА), г. Москва. E-mail: [zeninsergei@mail.ru](mailto:zeninsergei@mail.ru).

**Кутейников Дмитрий Леонидович** – кандидат юридических наук, старший преподаватель кафедры конституционного и муниципального права, Университет имени О. Е. Кутафина (МГЮА), г. Москва. E-mail: [kuteynikov@me.com](mailto:kuteynikov@me.com).

**Япрынцеv Иван Михайлович** – кандидат юридических наук, советник судьи, Конституционный Суд Российской Федерации, г. Москва. E-mail: [thevanfirst@gmail.com](mailto:thevanfirst@gmail.com).

**Ижаев Осман Аликович** – кандидат юридических наук, консультант юридического отдела, Московская городская избирательная комиссия, г. Москва. E-mail: [izhaev.osman@gmail.com](mailto:izhaev.osman@gmail.com).

*Статья поступила в редакцию 23 июня 2020 г.*

DOI: 10.14529/law200311

## **NATURAL LANGUAGE PROCESSING IN THE LAW-MAKING PROCESS**

**S. S. Zenin, D. L. Kuteinikov**

*University named after O. E. Kutafin (MSAL), Moscow, Russian Federation,*

**I. M. Yaprntsev**

*The constitutional Court of the Russian Federation, Moscow, Russian Federation,*

**O. A. Izhaev**

*Moscow city election Commission, Moscow, Russian Federation*

The article outlines the relationship between the introduction of machine learning technology into the legislative process with the need for natural language processing (NLP), which is a necessary element in this process.

The formation of databases covering the existing normative array is associated with the need to translate it into a machine-readable form, understandable to the algorithm. Using NLP allows performing such a transformation.

The authors describe examples of the use of NLP in the legal field. Taking into account the results of such use, the article assesses the potential for the implementation of NLP in the framework of legislative activity. At the same time, the specificity of legal texts is noted, which predetermines the need for additional requirements for the process of their processing.

In conclusion, the authors note some risks associated with the introduction of machine learning into the legislative process, but which can be minimized by using NLP.

**Keywords:** *law-making process, automatization, algorithm, machine learning, natural language processing, semantic meaning of the text.*

**Sergey Sergeevich Zenin** – Candidate of Sciences (Law), Director, Research Institute, University named after O. E. Kutafin (MSAL), Moscow, Russian Federation. E-mail: zeninsergei@mail.ru.

**Dmitry Leonidovich Kuteynikov** – Candidate of Sciences (Law), senior lecturer of the Department of Constitutional and Municipal Law, University named after O. E. Kutafin (MSAL), Moscow, Russian Federation. E-mail: kuteynikov@me.com.

**Ivan Mikhailovich Yapyryntsev** – Candidate of Sciences (Law), Advisor to the judge, Constitutional Court of the Russian Federation, Moscow, Russian Federation. E-mail: thevanfirst@gmail.com.

**Osman Alikovich Izhaev** – Candidate of Sciences (Law), consultant of the legal Department, Moscow city election Commission, Moscow, Russian Federation. E-mail: izhaev.osman@gmail.com.

*Received 23 June 2020.*

---

**ОБРАЗЕЦ ЦИТИРОВАНИЯ**

Зенин, С. С. Технология обработки естественного языка (NLP) в законодательном процессе / С. С. Зенин, Д. Л. Кутейников, И. М. Япрынцева, О. А. Ижаев // Вестник ЮУрГУ. Серия «Право». – 2020. – Т. 20, № 3. – С. 76–81. DOI: 10.14529/law200311.

**FOR CITATION**

Zenin S. S., Kuteinikov D. L., Yapyryntsev I. M., Izhaev O. A. Natural language processing in the law-making process. *Bulletin of the South Ural State University. Ser. Law*, 2020, vol. 20, no. 3, pp. 76–81. (in Russ.) DOI: 10.14529/law200311.

---