

А.О. Козлова, В.А. Ковязин, И.Р. Зулькарнеев

Тюменский государственный университет, г.Тюмень

УДК 004.056

АННОТИРОВАНИЕ НОРМАТИВНЫХ ПРАВОВЫХ АКТОВ РОССИЙСКОЙ ФЕДЕРАЦИИ В ОБЛАСТИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

Аннотация. В статье рассматривается инженерный подход извлечения информации из нормативных правовых актов Российской Федерации для определения списка обязательных требований и последующей их категоризации по принадлежности к информационной системе. Особое внимание уделяется оценке эффективности разработанной системы извлечения информации для выявления ее недостатков и составления этапов дальнейшего развития.

Ключевые слова: извлечение информации, обработка естественного языка, правовое обеспечение информационной безопасности, требования к защите информации.

Настоящее время характеризуется процессом информатизации общества, в основе которого лежит сбор, обработка, хранение, передача и использование информации. В связи с чем появилась возможность доступа к большому объему данных. Одновременно с этим сформировалась проблема защиты информации от постороннего доступа и нежелательного воздействия на нее.

Для решения вышеизложенной задачи права и обязанности обладателей информации и операторов информационной системы регулируются государством нормативными правовыми актами Российской Федерации, что включает законодательные и методические документы,

международные и отечественные стандарты, приказы и распоряжения, принимаемые руководителями различных организаций, которые постоянно совершенствуются и обновляются.

Многочисленность законодательных документов различной отраслевой направленности в области информационной безопасности показала, что возникает проблема определения перечня обязательных требований со ссылками на нормативные правовые акты по обеспечению защиты информации.

Поиск подходящих решений для составления перечня показал, что не существует системы, которая бы в нормативных правовых актах Российской Федерации точно определяла требования в области информационной безопасности. Отечественные исследования [1] велись для предсказания правовой классификации фактов, высказанных на естественном языке на примере Уголовного и Гражданского Кодекса Российской Федерации. Зарубежные разработки [2], [3] ориентированы на извлечение заголовков и подзаголовков разделов, расстояния и дат, именованных объектов в правовых и нормативных текстах, написанных на английском языке. В связи с этим было принято решение создать систему для автоматического извлечения информации о необходимых требованиях в нормативных правовых актах Российской Федерации по защите информации и преобразования их в структурированную аннотированную форму.

В ходе изучения [4] был выявлен инженерный подход для реализации системы, который опирается на то, что извлекаемая информация используется в рамках определённых языковых конструкций – лингвистических шаблонов, поэтому следующей нашей задачей было определение, как в нормативных правовых актах устанавливается требование.

Согласно стандарту [5] обязательное требование должно содержать в своей формулировке определенные слова, такие как "обязаны", "должны" и аналогичные по смыслу, а также исключать слова «рекомендуется», «по возможности», «вправе» и подобные. Данный факт позволил выделить характерные шаблоны в формулировках обязательных требований. Стандарт не определяет полный и исчерпывающий список слов, которые должны содержаться в формулировке требования, поэтому мы самостоятельно проанализировали нормативные правовые акты Российской Федерации в области информационной безопасности на предмет узконаправленных лингвистических шаблонов. В итоге был сформирован список необходимых для обработки документов конструкций:

- 1) обязаны;
- 2) запрещается + глагол;
- 3) запрещается распространение сведений/информации;
- 4) должны;
- 5) требования предусматривают;
- 6) не допускается;
- 7) осуществляется/определяется оператором в соответствии;
- 8) осуществляется/определяется в соответствии;
- 9) применяются средства защиты информации;
- 10) проводится с учетом гост;
- 11) реализацию требований;
- 12) соответствие требования;
- 13) выполнение ... требований.

Для списка информационных систем мы выделили ключевые слова, по которым определяли принадлежность требования к информационной системе. Выбран именно такой способ, потому что по тексту или названию документа в большинстве случаев встречаются сочетания слов,

подходящие только для определенной информационной системы, что позволяет ее идентифицировать.

Следующим шагом было создание системы, включающей в себя:

- 1) разбиение текстов нормативных правовых актов на структурные единицы для дальнейшей обработки (рис. 1);
- 2) приведение набора текстов к нормализованному виду при помощи методов NLP для устранения проблемы лингвистической изменчивости языка;

```
{
  "fz-149": {
    "name": "Об информации, информационных технологиях и о защите информации",
    "articles": {
      "статья 1.": {
        "1.": {
          "text": "1. Настоящий Федеральный закон...",
        },
        "2.": {
          "text": "2. Положения настоящего Федерального закона...",
        }
      }
    }
  }
}
```

Рис. 1. Структура обработанного нормативного правового акта.

- 3) выделение в нормализованных данных требований, соответствующих лингвистическим шаблонам;
- 4) определение, к какой информационной системе относятся выделенные требования;
- 5) аннотирование обрабатываемого корпуса данных (рис. 2).

```

{
  "fz-149": {
    "name": "Об информации, информационных технологиях и о защите информации",
    "articles": {
      "Статья 10.1.": {
        "1.": {
          "text": "1. Организатором распространения информации в сети \\"Интернет\\" является лицо...",
          "demand": false,
          "system": null
        },
        "2.": {
          "text": "2. Организатор распространения информации в сети \\"Интернет\\" обязан...",
          "demand": true,
          "system": "Информационная система общего пользования"
        },
        ...
      }
    }
  }
}

```

Рис. 2. Структура аннотированного нормативного правового акта.

Для оценки эффективности работы системы (таблица 1) использовались количественный метрики – точность и полнота, а также F-мера, позволяющая найти баланс между ними.

Таблица 1. Оценка эффективности системы.

№ проводимой оценки	Точность, %	Полнота, %	F-мера, %
1	67	78	72
2	82	93	87

В ходе первой оценки эффективность системы составила 72%, что мы посчитали слабым результатом, поэтому было проведено сравнение результатов экспертного и автоматизированного анализа нормативных правовых актов. По его итогам были исключены избыточные шаблоны, которые допускали большой процент ошибок, и добавлены новые более узкой направленности. После второй оценки эффективность системы достигла 87%.

Следующим шагом была оценка точности определения информационной системы, которая составила 90%.

На данном этапе была реализована система, которая предоставляет список требований со ссылками на нормативные правовые акты Российской Федерации по различным видам информационных систем и подготавливает аннотированный корпус для последующего изучения

правового обеспечения информационной безопасности. В дальнейшем планируется повысить эффективность системы с помощью внедрения методов машинного обучения и создания гибридной системы.

СПИСОК ЛИТЕРАТУРЫ

1. Крылов В.В., Крылов С.В, Жигалов Г.М. Построение языковой сети правового пространства для судебных приложений искусственного интеллекта. — сборник Шестой международной конференции «Актуальные проблемы системной и программной инженерии» — 2019 — № 2514 — с. 308-319 — [Электронный ресурс] — URL: <http://ceur-ws.org/Vol-2514/paper3.pdf> (дата обращения: 30.05.2020)
2. Opsomer B., De Meyer G., Cornelis C., Van Eetvelde G. Exploiting Properties of Legislative Texts to Improve Classification Accuracy. — Legal knowledge and information systems: JURIX 2009, the twenty-second annual conference — 2009 — Vol. 205 — p. 136-145 — [Электронный ресурс] — URL: <https://www.worldcat.org/title/legal-knowledge-and-information-systems-jurix-2009-the-twenty-second-annual-conference/oclc/861796265/viewport> (дата обращения: 30.05.2020)
3. Bommarito II M. J., Katz D.M., Detterman E.M. LexNLP: Natural language processing and information extraction for legal and regulatory texts. [Электронный ресурс] — URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3192101 (дата обращения: 30.05.2020)
4. Piskorski J., Yangarber R. Information Extraction: Past, Present and Future. — Multi-source, Multilingual Information Extraction and Summarization. Theory and Applications of Natural Language Processing — 2013 — p.23-49 — [Электронный ресурс] — URL:

<https://www.semanticscholar.org/paper/Chapter-2-Information-Extraction-%3A-Past-%2C-Present-Piskorski-Yangarber/59d1ec0e997d922004cb12e8399ba394c43aa72d> (дата обращения: 29.04.2020)

5. Стандарт качества нормативно-правового регулирования обязательных требований, утвержденный протоколом заседания проектного комитета от 24 апреля 2018 г. № 3. [Электронный ресурс] — URL: https://www.fsvps.ru/fsvps-docs/ru/practice/metod/standart_npa_reg.pdf (дата обращения: 21.05.2020)