

УДК 004.912:34

DOI 10.34822/1999-7604-2020-1-43-52

НАУЧНО-МЕТОДИЧЕСКИЕ ПОДХОДЫ К ПРОЦЕССУ ШАБЛОНИЗИРОВАНИЯ ДОКУМЕНТОВ ПРАВОВЫХ АКТОВ В ИНТЕГРИРОВАННОМ ПОЛНОТЕКСТОВОМ БАНКЕ ПРАВОВОЙ ИНФОРМАЦИИ

И. Ю. Баранов, С. В. Козлов, А. А. Невров ✉

Академия Федеральной службы охраны Российской Федерации, Орёл, Россия

✉ E-mail: newrow@mail.ru

В статье рассмотрены вопросы, касающиеся автоматизации преобразования форматов документов правовых актов на основе настраиваемых шаблонов документов. Правовые акты претерпевают изменения в ходе законодательной деятельности органов государственной власти, и задача поддержания актуальной редакции правового акта с увеличением количества редакций становится все более сложной. Часть правовых актов хранится в текстовом формате с устаревшим моноширинным форматированием и требует преобразования в современный формат, например «Документ Microsoft Word». Однако значительное число правовых актов не поддается автоматическому переформатированию. Особые сложности вызывает форматирование документов, содержащих таблицы, сформированные символами псевдографики, пробельными или иными разделителями. В статье представлены подходы к переформатированию таких документов на основе выделения структурных элементов документа и их преобразованию с использованием настраиваемых шаблонов. Результаты могут быть использованы в региональных и федеральных органах исполнительной власти при подготовке новых редакций правовых актов в современном формате «Документ Microsoft Word».

Ключевые слова: правовой акт, формат моноширинного текстового документа, преобразование форматов.

SCIENTIFIC AND METHODOLOGICAL APPROACHES TO LEGAL ACTS FORMATTING IN INTEGRATED DATABANK OF LEGAL INFORMATION

I. Yu. Baranov, S. V. Kozlov, A. A. Nevrov ✉

Academy of the Federal Guard Service of the Russian Federation, Oryol, Russia

✉ E-mail: newrow@mail.ru

The article considers automatic format conversion of legal acts based on custom document templates. Legal acts undergo changes as a result of activities of state authorities, and the task of maintaining the correct edition of a legal act with an increase in the number of revisions becomes more and more difficult. The part of legal acts is stored in an outdated monospace format, and require conversion into a modern format such as Microsoft Word Document. However, many of them cannot be converted automatically. Formatting documents with tables formed by pseudographic characters, whitespaces or other dividers cause a particular difficulty. The approaches to reformatting such documents based on the selection of the structural elements of the document and their transformation using custom templates are presented. The results can be used in regional and federal executive authorities in the preparation of new versions of legal acts in the modern format of Microsoft Word Document.

Keywords: legal act, monospaced text document, format conversion.

Введение. Интегрированный полнотекстовый банк правовой информации (далее – ИПБПИ) – эталонный банк данных правовой информации, содержащий правовые акты (далее – ПА), в силу сложившихся причин хранящиеся в устаревшем формате моноширинного текстового документа (далее – ФМТД), форматирование в котором производилось не специ-

альными управляющими тегами документа, а пробелами. В настоящее время число подобных документов в ИПБПИ составляет около 400 тыс. (1969–2000 гг.).

Моноширинные или непропорциональные шрифты были разработаны для пишущих машинок. За счет одинаковой ширины для всех символов шрифта точность выравнивания текста составляла один любой символ, включая пробел. Такие шрифты применяются для набора таблиц, листинга программ и в других случаях, когда требуется полное совпадение колонок и знаков по вертикали набора [1]. Для компьютеров с алфавитно-цифровыми устройствами ввода-вывода в наборы символов кроме букв алфавита, цифр, пробелов и знаков препинания входили символы псевдографики для представления таблиц и рисунков. К моноширинным шрифтам относятся: Courier, Courier New, Terminus, Monaco, Lusida Console и др. Используемое на рабочих местах программное обеспечение Microsoft Office предполагает открытие документов .txt и .rtf с ФМТД из ИПБПИ с применением моноширинного шрифта Courier New.

Недостатками моноширинных шрифтов для представления документов на бумаге и экране компьютеров являются [2]:

- нечитабельность букв – при одинаковой ширине места для одной буквы некоторые из них оптически выглядели трудноразличимыми (например, буквы «М», «Ш», «Ж»), особенно в полужирном исполнении;
- нечитабельность текста – при размещении каждой буквы на своей позиции образуются оптические разрывы слов, например, при размещении рядом с буквами «А», «Т» и др.;
- совпадение по начертанию буквы «О» и цифры «0» и др.

В бумажной типографике данные недостатки устранены за счет применения засечных моноширинных шрифтов, пропорциональных шрифтов, кернинга и т. д. Для компьютерных устройств возможность использования таких приемов появилась с повсеместным внедрением в рабочих станциях и персональных ЭВМ графической оболочки с отрисовкой любой буквы, символа, рисунка с помощью точек. Это позволило использовать рациональные пропорциональные шрифты в документах компьютерного исполнения, сделать их удобочитаемыми и повысить информационную емкость на страницу текста.

Правовые акты в ФМТД, размещенные в ИПБПИ, имеют ряд существенных недостатков, особенно остро проявляющихся при поддержании актуальной редакции документа. Под видом ПА в ИПБПИ будем понимать перечень видов актов, приведенных в документе по адресу: www.pravo.gov.ru/ips/?start_search&fattrib=1. Под формой ПА в ИПБПИ будем понимать вид ПА в формате, принятом в федеральном органе исполнительной власти (далее – ФОИВ) и региональном органе исполнительной власти (далее – РОИВ).

Перечень видов ПА в ИПБПИ представляет собой ограниченное множество, его расширение проводится редко. В свою очередь, формы ПА не приведены к общему стандарту, несмотря на существование руководящих документов, предусматривающих применение определенных правил оформления ПА [3]. Правила нарушаются по различным причинам. Вследствие этого изменение документов в ФМТД, требующееся при поступлении новых ПА, в формат, обеспечивающий возможность создания актуальной редакции изменяемого документа, не является детерминированным процессом и в настоящее время требует рутинной ручной обработки. Такая обработка входного потока ПА занимает большой ресурс рабочего времени и нуждается в автоматизации.

Структура текста правового акта. Для автоматизации процесса при использовании нежестко структурированного текста используется направление автоматического анализа вероятностного соотнесения ключевых слов и словосочетаний документа с заранее заданными словарными эталонами, привязанными к определенным стилям, предметным областям, подъязыкам или тематическим рубрикам. К таким ключевым словам относятся исходные атрибуты документов:

- реквизиты органа, издавшего ПА ФОИВ или РОИВ;
- вид ПА ФОИВ или РОИВ;
- место подписания ПА ФОИВ или РОИВ;

- дата подписания ПА ФОИВ или РОИВ;
- номер ПА ФОИВ или РОИВ;
- наименование ПА ФОИВ или РОИВ;
- реквизиты регистрации в Минюсте России;
- содержание ПА ФОИВ или РОИВ;
- подписант ПА ФОИВ или РОИВ (должность, инициалы, фамилия).

Обобщенный порядок конвертирования из ПА в ФМТД в ПА в пропорциональном шрифте с известной структурой представлен на рис. 1.

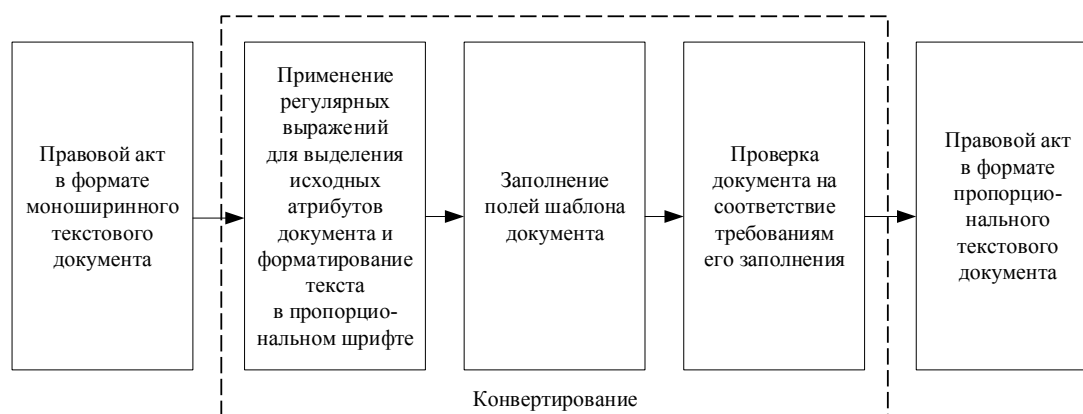


Рис. 1. Макроструктура обобщенного порядка конвертирования из ПА в ФМТД в ПА в пропорциональном шрифте

Примечание: составлено авторами.

В качестве примера на рис. 2–3 приведено представление одного вида ПА «Закон» РОИВ в различных формах в ФМТД, где исходные атрибуты документов размещены в разных местах.

27 апреля 1998 года	68-03
ЗАКОН АМУРСКОЙ ОБЛАСТИ	
О БЮДЖЕТЕ ТЕРРИТОРИАЛЬНОГО ДОРОЖНОГО ФОНДА АМУРСКОЙ ОБЛАСТИ НА 1998 ГОД	
Принят Амурским областным Советом народных депутатов 16 апреля 1998 года	
(В редакции Закона Амурской области от 02.10.98 г. N 93-03)	
Статья 1	
Утвердить бюджет территориального дорожного фонда области на 1998	
...	
Глава Администрации Амурской области	А.Н. БЕЛОНОВ
г. Благовещенск 27 апреля 1998 года N 68-03	

Рис. 2. Форма представления закона Амурской области [4]

РОССИЙСКАЯ ФЕДЕРАЦИЯ ЗАКОН ЧЕЛЯБИНСКОЙ ОБЛАСТИ	
О бюджете областного дорожного фонда на 1998 год	
Статья 1. Утвердить бюджет областного дорожного фонда на 1998 год по доходам в сумме 1556890 тыс. руб. и по расходам в сумме 1556890 тыс. руб.	
...	
Губернатор Челябинской области N 40-30 от 16 апреля 1998 года	П.И.Сумин 30.04.98 г.

Рис. 3. Форма представления закона Челябинской области [5]

Наибольшую трудность при конвертировании ПА в ФМТД представляет работа с таблицами и рисунками, выполненными символами псевдографики. При этом оформление таблиц ПА в ФМТД из ИПБПИ не стандартизовано и выполнялось в ФОИВ различными способами, что сильно затрудняет выявление общих табличных признаков (заголовка, тела таблицы, объединения ячеек и др.) для последующего создания шаблонов. Примеры таблиц ПА в ФМТД из ИПБПИ показаны на рис. 4–6.

П Е Р Е Ч Е Н Ь продукции военного назначения, в отношении которой открытому акционерному обществу "Научно-производственное предприятие "Звезда" предоставляется право осуществления внешнеторговой деятельности	
Номенклатура продукции военного назначения	Виды работ и услуг
Катапультиные кресла К-36, К-37 и их модификации	проведение научно-исследовательских и опытно-конструкторских работ поставка на экспорт
Амортизационные кресла типа "Памир", "Казбек" и их модификации	поставка на экспорт комплектующих и запасных частей оказание технического содействия в

**Рис. 4. Таблица из Распоряжения Правительства Российской Федерации
№ 1849-р от 31 декабря 1997 г. [6]**

Размер облагаемого совокупного дохода, полученного в календарном году	Сумма налога
до 20 000 руб.	12 процентов
от 20 001 до 40 000 руб.	2 400 руб. + 15 процентов с суммы, превышающей 20 000 руб.

**Рис. 5. Пример таблицы из Федерального закона
№ 159-ФЗ от 31 декабря 1997 г. [7]**

Выданы водительские удостоверения				
Дата выдачи	Серия, N водитель- ского удосто- верения	Разрешенные категории	Наименование подразделения, выдавшего удостоверение	Подпись должностного лица, печать

Рис. 6. Приказ Министерства внутренних дел Российской Федерации № 860 от 30 декабря 1997 г. [8]

Формирование шаблона правового акта. В связи с разными формами ПА, принятыми в ФОИВ и РОИВ, целесообразна разработка создаваемых и редактируемых пользователем шаблонов документов.

Для решения задачи автоматизации по созданию шаблонов и конвертированию документов из ФМТД необходимо отметить общность форматов входного и выходного потоков ПА для проектируемого средства автоматизации. В качестве такого формата в системе ИПБПИ принят проприетарный межплатформенный формат хранения форматированного текста – Rich Text Format (rtf) [9].

В случаях конвертирования ПА в ФМТД в формат документов с пропорциональным шрифтом, но без изменения его структуры требуется разработка инструмента, позволяющего пользователю определять структуру исходного документа, извлекать исходные атрибуты документа, создавать шаблон и помещать исходные атрибуты в поля шаблона (рис. 7).

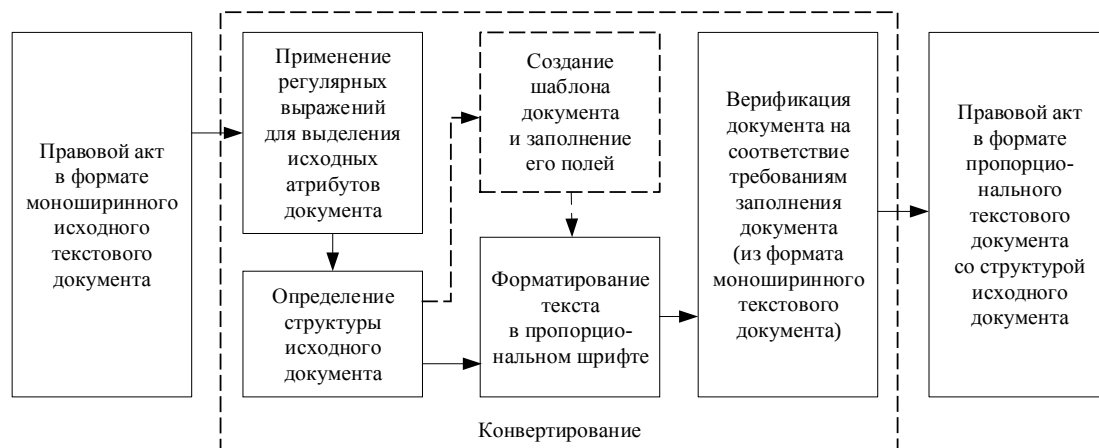


Рис. 7. Макроструктура порядка конвертирования из ПА в ФМТД в ПА в пропорциональном шрифте с сохранением структуры документа

Примечание: составлено авторами.

Если принять каждое поле исходных атрибутов документа как блок, то задача разработки структуры для создания шаблона сводится к взаимному расположению этих блоков в соответствии с их размещением в документе в ФМТД. Таким образом, под структурой документа будем понимать взаимное размещение элементов исходных атрибутов документа на странице. Созданный и верифицированный шаблон вида документа сохраняется в базе шаблонов и в дальнейшем применяется, минуя этапы определения структуры и создания шаблона.

Структурирование правового акта. Первичным этапом, предшествующим определению структуры документа, является применение регулярных выражений [10] (далее – РВ) для выделения исходных атрибутов документа. Использование РВ на сегодняшний день яв-

ляется наиболее развитым и проработанным научно-методологическим подходом для поиска и замены в тексте [11]. Ключевым на данном этапе является составление РВ, позволяющих найти в тексте ПА требуемый атрибут документа. РВ представляют собой формальный язык поиска и осуществления манипуляций с подстроками в тексте, который основан на использовании метасимволов. Следовательно, для составления требуемых РВ необходимо знание этого языка. В качестве средств поддержки создания РВ могут выступать существующие конструкторы РВ, например [12]. Ниже приведен пример РВ, позволяющего найти атрибут «Дата издания ПА» в областном законе Челябинской области (см. рис. 3):

от (\d+) (января|февраля|марта|апреля|мая|июня|июля|августа|
(?:сентя|октя|ноя|дека)бря) ([0-9]{4}) года

Найденные атрибуты документа позволяют определить его структуру, то есть найти позиции символов (номера строк, столбцов) в исходном документе в ФМТД, относящиеся к определенным атрибутам документа. Строки исходного документа, не отнесенные к атрибутам ПА, помечаются как содержимое ПА.

На этапе создания шаблона происходит определение необходимых и опциональных атрибутов документа. Для каждого атрибута составляется РВ, позволяющее найти этот атрибут в тексте ПА. Затем для каждого атрибута составляются правила его размещения и оформления в тексте преобразованного ПА. Размещение атрибута определяется его положением относительно других атрибутов (очередность в документе), а также видом горизонтального выравнивания в строке.

Для описания правил поиска и размещения атрибутов целесообразным является использование языка XML [13]. Этот язык сочетает простоту описания и гибкость, позволяя расширять или менять состав тегов и атрибутов. Кроме того, синтаксис языка дает возможность, используя распространенные правила именования тегов, сохранять в документе семантику, понятную разработчикам программного обеспечения. В качестве примера возможного описания шаблона документа ниже приведен фрагмент XML-файла, описывающего правовой акт, выпущенный администрацией Костромской области:

```
<?xml version="1.1" encoding="UTF-8" ?>
<template title="Шаблон Кострома-1">
<field name="type" desc="Вид правового акта">
<input color="#206020" regex="(постановление|распоряжение)"/>
<output order="30" align="center" capitalize="true" format="\n$1\n" style=""/>
</field>
<field name="number" desc="Номер">
<input color="#602080" regex="N (\d+)"/>
<output format="№ $1" align="left" style="font-weight:bold" />
</field>
<field name="date" desc="Дата">
<input color="#206080" regex="от (\d+) (января|февраля|марта|апреля|мая|июня|июля|
августа|(?::сентя|октя|ноя|дека)бря) ([0-9]{4}) года"/>
<output format="от $1 $2 $3 года" />
</field>
...
```

Нормализация текста содержимого правового акта. Содержимое ПА при конвертировании в современный формат требует ряда преобразований, обусловленных тем, что текст в ФМТД содержит пробелы, знаки табуляции и разрыва строки, являющиеся управляющими символами в современном формате. Процесс конвертирования включает разбиение на параграфы, нормализацию параграфа, форматирование параграфа в соответствии с его форматом в исходном документе.

Разбиение на параграфы основывается на поиске начала и конца блока текста, который относится к одному параграфу. Поиск начала и конца блока основывается на поиске красной строки – строки, начинающейся с задаваемого числа пробелов. Концом блока является пустая строка и строка, имеющая выравнивание, отличное от предыдущей. В настраиваемом шаблоне документа могут указываться параметры форматирования исходного документа: ширина строки (по умолчанию 70 символов), отступ красной строки (по умолчанию 5 пробелов).

Нормализация параграфа заключается в слиянии строк с объединением слов, разделенных переносом, удалении пробелов в начале и конце параграфа и удалении кратных пробелов, применяемых в ФМТД для выравнивания текста ПА по ширине. В настоящее время этот процесс не всегда безошибочен. Когда на границе строки оказывается перенос слова, обозначаемый в ФМТД (например, знаком «дефис»), он должен быть удален, а слово, оказавшееся перенесенным, должно быть объединено. Однако в ПА встречаются ситуации, когда присутствует дефис в конце строки внутри сложносоставного слова. В таких случаях процедура удаления дефиса и объединения слова является некорректной. Возможный подход к устранению такого рода ошибок – ведение словаря сложносоставных слов и проверка перенесенных слов перед объединением по этому словарю.

Конвертирование таблиц, содержащихся в документе. Конвертирование ПА в ФМТД, содержащего таблицы, должно предваряться заданием вертикальных и горизонтальных разделителей ячеек таблицы. Они могут различаться в зависимости от вида ПА (рис. 4–6) и региона, в котором принят ПА. В настраиваемом шаблоне предусмотрены поля для задания горизонтальных и вертикальных разделителей. В качестве горизонтальных разделителей (отделяют друг от друга строки) в таблицах, содержащихся в ПА, как правило, применяются знаки «минус», «дефис», «тире» и подобные сходные по начертанию символы. В связи с тем, что в моноширинных шрифтах зачастую сложно или невозможно на глаз определить, какой из указанных символов используется, при конвертировании таблиц предлагается копировать их в буфер обмена из текста ПА и вставлять в соответствующие поля шаблона, предназначенные для задания разделителей в настраиваемом шаблоне.

Конвертирование таблицы начинается с выделения области документа, содержащего таблицу. Затем в выделенной области ищутся горизонтальные и вертикальные разделители, определяются количество строк и столбцов, позиции их начала. В новом документе создается пустая таблица с нужным количеством столбцов и строк. На основе найденных позиций выделяются ячейки таблицы. Содержащийся в ячейках текст копируется в ячейки новой таблицы, при этом текст каждой ячейки нормализуется по правилам нормализации параграфа. Данный этап преобразования таблицы хорошо себя зарекомендовал на правильно отформатированных исходных таблицах (рис. 6). Для таких таблиц возможно осуществить автоматическое преобразование.

Эвристический подход к нахождению границ ячеек. В значительном числе документов присутствуют таблицы, в которых разделение на столбцы и строки осуществлено неявно (рис. 4–5). Для таких случаев разработана процедура эвристического поиска вертикальных и горизонтальных границ в таблице. Эвристический поиск вертикальных границ столбцов в таблице базируется на предположениях, что левая граница столбца должна содержать начало слов (рис. 5) и слова в этой позиции не должны «разделяться». Процедура поиска заключается в нахождении в тексте позиций, в которых начинаются слова. Поиск ведется с использованием регулярного выражения «\s\S», которому соответствуют пары символов, первый из которых является пробельным, а второй – непробельным. Каждая найденная позиция увеличивает счетчик начала слов для нее. Затем каждая позиция проверяется на предмет того, разделяет она какое-нибудь слово в таблице или нет. Проверка осуществляется с помощью регулярного выражения «\S\S». Этому регулярному выражению соответствуют пары символов, каждый из которых является непробельным. Если есть разделение, то происходит уменьшение счетчика в этой позиции.

Пусть T – матрица, каждый элемент $t_{i,j}$ которой представляет собой символ анализируемого участка текста в i -ой строке в j -ой позиции.

Формула для расчета коэффициента наличия вертикальной границы в таблице в позиции c -го столбца $V(c)$ приведена в (1):

$$V(c) = \sum_{r=0}^{N-1} \frac{B(r, c)}{N}, \quad (1)$$

где r – номер строки;

c – номер столбца;

N – число строк в анализируемом тексте;

$B(r, c)$ – функция принадлежности позиции вертикальной границе в таблице, вычисляемая по формуле (2):

$$B(r, c) = \begin{cases} 1, & t_{r,c} + t_{r,c+1} \in G, \\ -1, & t_{r,c} + t_{r,c+1} \in W, \\ 0, & \text{иначе,} \end{cases} \quad (2)$$

где двумерный массив T – анализируемый участок текста из ПА в ФМТД;

операция $t_{r,c} + t_{r,c+1}$ означает конкатенацию двух соседних символов;

G – множество пар символов, задаваемое регулярным выражением «\s\S»;

W – множество пар символов, задаваемое регулярным выражением «\S\S».

Получаемый коэффициент $V(c)$ (1) содержит значения, характеризующие степень ожидания наличия вертикальной границы в столбце c . Для принятия решения о задании границы в таблице данный коэффициент можно сравнить с пороговым значением T_v , задаваемым в настраиваемом шаблоне. По умолчанию пороговое значение равно 0,5.

Эвристический поиск горизонтальных границ в таблице заключается в нахождении строк текста, в которых количество непробельных символов значительно превышает их количество в предыдущей строке. Для нахождения таких строк текста вводится коэффициент $H(r)$, характеризующий «наполненность» строки по сравнению с предыдущей.

$$H(r) = \frac{S(r) - S(r-1)}{M}, \quad (3)$$

где $S(r)$ – функция, возвращающая количество непробельных символов в строке с индексом r ;

M – максимальная длина строки анализируемого текста.

Для принятия решения о наличии горизонтальной границы в строке с индексом r значение полученного коэффициента $H(r)$ (3) сравнивается с пороговым значением Th , задаваемым в настраиваемом шаблоне. По умолчанию значение Th равно 0,15.

Автоматизированный режим конвертирования таблиц. Применяемые эвристики позволяют определить границы строк и столбцов в таблицах с неявно заданными ячейками и автоматизировать их конвертирование. Однако в ряде существующих ПА присутствуют таблицы, на которых данные эвристики не срабатывают из-за ошибочного оформления таблиц (например, если размеры столбцов в заголовочной части таблицы не совпадают с размерами, определенными эвристическими методами). В этих случаях оператору, осуществляющему конвертирование ПА, следует предоставить инструментарий для автоматизированного конвертирования таблиц с визуальным контролем промежуточных результатов разбиения таблицы.

Авторами предложен механизм преобразования текста в таблицу на основе представления текста ПА в виде двумерной таблицы, элементы которой содержат один символ исходного текста ПА в ФМТД и сведения о границе, которая проходит через этот элемент. Элемент может не содержать границ, иметь горизонтальную верхнюю, горизонтальную нижнюю, вертикальную левую, вертикальную правую границы или их комбинации.

Разработанный авторами инструментарий позволяет добавлять и удалять границы в таблице, перемещать блоки (прямоугольные участки) текста относительно других блоков, добавлять и удалять блоки текста. Кроме того, для удобства оператора предусмотрена возможность отмены действий, касающихся редактирования ячеек таблицы.

В случае работы в визуальном режиме конвертирования таблицы оператору предоставляется возможность провести ряд автоматических проверок корректности задания границ таблицы. Одной из таких проверок является проверка отсутствия «разрывов» в границе таблицы. Она осуществляется на базе разработанного авторами алгоритма, который состоит в последовательном просмотре каждого элемента матрицы и поиске горизонтальных или вертикальных «разрывов» путем достаточно сложной проверки типов границ для соседних с просматриваемым элементом. При обнаружении единичного разрыва он устраняется автоматически. Разрывы длиной более одного элемента подсвечиваются красным цветом, сигнализируя оператору о возможной ошибке.

Авторами разработан алгоритм, позволяющий проверить корректность задания границ таблицы на основе предположения о том, что каждая ячейка таблицы должна быть прямоугольной. Разработанный алгоритм основан на алгоритме поиска в ширину и заключается в нахождении описывающего ячейку таблицы прямоугольника и проверке вхождения всех элементов ячейки таблицы, не имеющих границ, в эту прямоугольную область. Если множества элементов не совпадают, то такая ячейка подсвечивается красным цветом, сигнализируя оператору о возможной ошибке.

Подходы к конвертированию рисунков, содержащихся в правовых актах. Некоторая часть ПА в ФМТД содержит рисунки, сформированные символами псевдографики. Чаще всего они содержат блок-схемы каких-либо процессов и несут значительную семантическую нагрузку. Для их конвертирования должны быть сформулированы новые подходы и разработаны новые алгоритмы, потому что разработанные авторами подходы к конвертированию таблиц напрямую не применимы к конвертированию данных рисунков. Эта разработка является направлением для дальнейших исследований.

Выводы:

1. В связи с разнообразием подходов в федеральных органах исполнительной власти и региональных органах исполнительной власти к структуре документов ПА разработка универсального шаблона для их конвертирования нецелесообразна.
2. Автоматизация процесса конвертирования (рис. 7) должна включать в себя в том числе процесс определения структуры исходного документа без участия пользователя или с минимальными затратами ручного труда.
3. Определение структуры исходного документа и создание его шаблона выполняются один раз для группы однотипных видов правовых актов федеральных органов исполнительной власти или региональных органов исполнительной власти.
4. Структурирование ПА выполняется на основе регулярных выражений, задаваемых в шаблоне. Для каждого структурного элемента задаются правила его форматирования в выходном документе.
5. Часть ПА, не отнесенная к структурным элементам, является его содержимым. Содержимое ПА подлежит нормализации, осуществляемой на основе ряда параметров, задаваемых в шаблоне документа.
6. Значительную сложность представляет конвертирование таблиц, содержащихся в ПА. Этот процесс основывается на разбиении текста на ячейки с использованием настраиваемых разделителей или эвристик. Для сложных случаев предусмотрен автоматизированный режим конвертирования с возможностью визуального контроля и правки содержимого и границ таблицы.
7. Направлением для дальнейших исследований является поиск подходов к конвертированию рисунков, задаваемых символами псевдографики.

Литература

1. Моноширинные шрифты / Справ. компании Paratype. URL: <http://rus.paratype.ru/catalog-chapter/m-1dfd7f9c#monospaced-fonts> (дата обращения: 26.11.2019).
2. Королькова А. Живая типографика. М. : IndexMarket, 2012. 224 с.
3. Чуковенков А. Ю. Юридическая техника и правила оформления документов // Секретарь-референт. 2009. № 5. С. 17–21.
4. О бюджете территориального дорожного фонда Амурской области на 1998 год : закон Амурской области от 27 апр. 1998 г. : принят Амурским областным Советом народных депутатов 16 апреля 1998 г. // URL: http://www.pravo.gov.ru/proxy/ips/?docbody=&link_id=5&nd=181008574 (дата обращения: 09.12.2019).
5. О бюджете областного дорожного фонда на 1998 год : закон Челябинской области от 16 апр. 1998 г. // URL: http://www.pravo.gov.ru/proxy/ips/?docbody=&link_id=5&nd=162022201 (дата обращения: 09.12.2019).
6. Распоряжение Российской Федерации № 1849-р от 31 дек. 1997 г. // URL: http://www.pravo.gov.ru/proxy/ips/?docbody=&link_id=12&nd=102050902 (дата обращения: 09.12.2019).
7. О внесении изменений и дополнений в Закон Российской Федерации «О подоходном налоге с физических лиц» : федер. закон Рос. Федерации от 31 дек. 1997 г. № 159-ФЗ : принят Гос. Думой Федер. Собр. Рос. Федерации 19 декабря 1997 г. : одобр. Советом Федерации Федер. Собр. Рос. Федерации 24 декабря 1997 г. // URL: http://www.pravo.gov.ru/proxy/ips/?docbody=&link_id=2&nd=102050932 (дата обращения: 09.12.2019).
8. Об утверждении Инструкции по применению Правил сдачи квалификационных экзаменов и выдачи водительских удостоверений : приказ Министерства внутренних дел Российской Федерации от 30 дек. 1997 г. № 860 // URL: http://www.pravo.gov.ru/proxy/ips/?docbody=&link_id=12&nd=102093565 (дата обращения: 09.12.2019).
9. Rich Text Format Specification. URL: <https://www.microsoft.com/en-us/download/confirmation.aspx?id=10725&6B49FDFB-8E5B-4B07-BC31-15695C5A2143=1> (дата обращения: 02.12.2019).
10. Фицджеральд М. Регулярные выражения: основы. М. : ИД «Вильямс», 2015. 144 с.
11. Хопкрофт Д., Мотвани Р., Ульман Дж. Введение в теорию автоматов, языков и вычислений. М. : ИД «Вильямс», 2008. 529 с.
12. Конструктор регулярных выражений. URL: <https://easyregex.ru/constructor> (дата обращения: 29.01.2020).
13. Хантер Д., Рафтер Дж. XML. Базовый курс. 4-е издание. М. : ИД «Вильямс», 2018. 1344 с.