

# Исследование алгоритмов анализа содержимого технической документации

А. С. Сергеев, e-mail: toljaserg7@mail.ru

Московский Авиационный Институт (Национальный  
Исследовательский Университет)

***Аннотация.** На основе требований ГОСТ Р 7.0.100-2018 разработаны методы определения типа цитируемого ресурса и определения корректности библиографических ссылок и библиографических списков в научно-технической документации.*

***Ключевые слова:** цитирование, библиография, ссылки, источники, библиографическое описание, документация.*

## Введение

При разработке и поддержке научно-технической документации, авторам приходится уделять значительное количество своего времени не содержательной части работы, а оформлению документа согласно требованиям государственных стандартов, конкретных компаний, фирм, учебных учреждений, научных учреждений и т.п. И если корректное оформление шрифта, абзацев рисунков, диаграмм и пр. обычно не вызывает больших сложностей, то задачи по формированию списка аббревиатур или библиографического списка могут требовать не только значительных временных затрат, но и достаточной компетенции в данном вопросе.

Существует великое множество стилей оформления научно-технической документации, будь то Chicago, Harvard, LSNS или стиль ГОСТ РФ, и все они предъявляют различные требования к оформлению библиографических списков. Ввиду такого разнообразия, авторам, для обеспечения корректности оформления библиографического списка необходимо изучать официальную документацию, регламентирующую его оформление, однако зачастую авторы используют в т.ч. альтернативные источники информации, например, ресурсы в сети Интернет, которые не всегда корректно передают требования первоисточника. Западные стили цитирования, например, вышеупомянутые Chicago и Harvard, в большинстве своём имеют гораздо меньше обязательных для библиографической ссылки элементов, благодаря чему автоматизировать работу с ними гораздо проще, когда как для отечественных государственных стандартов практически нет подобных решений.

На основании вышеизложенного актуальна задача разработки методов валидации библиографических ссылок и списков, с помощью которых можно было бы проверять корректность цитирования. Это может оказать положительное влияние на качество формируемых библиографических списков, а главное – способствовать экономии времени авторов, чтобы они могли уделить больше времени содержательной части работы.

### **1. Анализ типов цитируемых ресурсов и требований к ним**

Ввиду того, что требования к оформлению библиографических ссылок в рамках одного стиля различаются между собой ввиду разных типов цитируемых ресурсов, необходимо рассмотреть классификацию библиографических ссылок в документе по типам. Таким образом, необходимо отделить библиографическое описание статьи от, например, библиографического описания книги. Для такой классификации библиографических ссылок были разработаны два метода: прямой и гибридный.

1. Прямой метод – метод, согласно которому тип цитируемого ресурса определяется в результате нахождения в тексте библиографической ссылки ключевых слов, определённого порядка расстановки знаков препинания, либо сегментов библиографического описания, присущих исключительно этому типу цитируемого ресурса. Данный метод применим к таким типам цитируемых ресурсов, как:

- Диссертация либо автореферат диссертации. Определяется по словам «диссертация на соискание...» либо «автореферат диссертации на соискание...»;
- Законодательные материалы. Определяется по словам «принят \*\*\*\*\*; одобрен \*\*\*\*\*» или «утверждён \*\*\*\*\*»;
- Государственные Стандарты (ГОСТ). Определяются по наличию аббревиатуры «ГОСТ» в начале библиографической ссылки, словам «национальный стандарт Российской Федерации»;
- Патентные документы. Определяются по слову «Патент №» в начале библиографического описания;
- Газеты. Определяются по наличию последовательности символов «газета /»;
- Ресурсы в сети Интернет (сайты). Определяются по наличию последовательности символов «: сайт. ->» сразу после области заглавия.

2. Гибридный метод – метод, согласно которому тип цитируемого ресурса определяется не только в результате нахождения в тексте библиографической ссылки ключевых слов, определённого порядка

расстановки знаков препинания, либо сегментов библиографического описания, присущих исключительно этому типу цитируемого ресурса, но и в результате проверки на отсутствие в тексте библиографического описания сегментов, данному типу цитируемого ресурса не присущих. Данный метод необходим для выделения из библиографического списка ссылок, тип которых невозможно определить прямым методом. К такому типу цитируемых ресурсов относятся статьи в научных журналах. Для соответствия данному типу цитируемого ресурса, библиографическая ссылка должна:

- Содержать последовательность символов «//» - признак многоуровневого библиографического описания;
- Содержать диапазон страниц журнала в конце библиографического описания, шаблон: «. - \*-\* с.» или «. – с. \*- \*»;

Не содержать следующего текста:

- «автореферат диссертации»
- «диссертация»
- «изд.», «Издание», «испр. и доп.» «исправленное и дополненное» и прочее, относящееся к книжным изданиям
- «ГОСТ», «Патент №» в начале библиографического описания
- «одобрен», «утверждён»
- «: сайт. -» сразу после заглавия
- Номера ISBN.

Методы применимы для наиболее популярных типов цитируемых ресурсов, так как используемые в ГОСТ редкие типы, например, «CD-диски» в настоящий момент практически не цитируются.

Основная сложность задачи классификации библиографических описаний заключается в том, что, например, в рамках требований ГОСТ Р 7.0.100-2018 области и сегменты библиографического описания подразделяются не только на «обязательные» и «факультативные», но и «условно-обязательные». Условно обязательный сегмент должен фигурировать в тексте библиографического описания, если информация, которая должна в нём содержаться, есть в наличии. Например, номера ISBN у книги иногда (редко) может и не быть, но во всех случаях, когда он есть, он обязателен для упоминания в библиографическом описании. Однако, авторы библиографических описаний нередко пользуются этой «условно обязательностью» и не добавляют такие сегменты библиографического описания вне зависимости от наличия нужной информации, и чисто технически это не является ошибкой в оформлении.

Таким образом, ни один метод классификации библиографического описания не может обеспечить 100% точность определения типа цитируемого ресурса, однако вышеописанные методы обеспечивают необходимую точность для дальнейшей валидации библиографических ссылок.

## 2. Валидация библиографического описания

Для валидации библиографического описания необходимо составить шаблоны идеального описания для каждого типа цитируемого ресурса с учётом как обязательных, так и условно-обязательных элементов. Валидация факультативных элементов в большинстве случаев не представляется возможной, так как обычно строгого регламента по содержанию факультативных элементов нет, к тому же они довольно редко фигурируют в библиографическом описании на практике.

Элементы библиографического описания, соответственно, также надо классифицировать согласно ГОСТ Р 7.0.100-2018 для корректной валидации. Если нам известен тип цитируемого ресурса – «ссылка» (например, в результате классификации прямым методом), классифицировать области не является сложной задачей, так как в рамках ГОСТа элементы и области отделяются друг от друга конкретными последовательностями знаков препинания, они описаны на таблице 1:

Таблица 1. Предписанные знаки элементов библиографического описания

Области описания	Предписанный знак	Элемент описания
Область заглавия и сведений об ответственности	=	Основное заглавие
	;	Параллельное заглавие
	/ ;	Сведения, относящиеся к заглавию
Область издания		Сведения об ответственности (первые   последующие)
	.-	Сведения об издании
Область публикации,	,	Дополнительные сведения об издании
	.-	Место публикации

производства, распространения, пр.	:	Имя издателя, производителя, распространителя
	,	Дата публикации
Область физической характеристики	.-	Сведения об объёме
Область примечания	.-	Сведения об электронном адресе
	.-	Остальные примечания
Область идентификатора	.-	Идентификатор ресурса
Область вида содержания и средства доступа	-   .	Вид содержания (первый   последующие)
	(начинается со строчной буквы)	Характеристика содержания
	:	Средство доступа

Стоит также отметить, что далеко не все элементы библиографического описания возможно валидировать. В основном проверка корректности может быть осуществлена только для «технических» элементов, таких как количество страниц, ISBN/ISSN номер, DOI и т.п., однако и в оформлении библиографического описания основная сложность заключается не в составлении содержания его элементов, а как раз в расстановке знаков препинания между ними. А проверять корректность расставления знаков препинания согласно государственному стандарту возможно с помощью программных средств.

### **Заключение**

Описанные в статье методы классификации библиографического описания и его дальнейшей валидации позволяют облегчить работу по подготовке научно-технической документации, также значительно повысить качество библиографического описания и его соответствие требованиям стиля, в котором оформляется документ.

Методы классификации учитывают разнообразие типов цитируемых ресурсов и наличие нерегламентированных по содержанию элементов библиографического описания, кроме того, включают в также стратегию «от обратного», когда тип цитируемого ресурса ссылкой определяется в том числе по отсутствию отдельных элементов.

Выделенные из ГОСТ Р 7.0.100-2018 требования к знакам препинания между областями и элементами библиографического описания позволяют не только определять типы элементов с достаточной точностью, но и проводить валидацию самих последовательностей знаков препинания.

### **Литература**

1. ГОСТ Р 7.0.100-2018. Система стандартов по информации, библиотечному и издательскому делу. Библиографическая запись. Библиографическое описание. Общие требования и правила составления. Утвержден и введен в действие Приказом федерального агентства по техническому регулированию и метрологии от 03.12.2018 №1050-ст – URL: <http://docs.cntd.ru/document/12000161674> (дата обращения: 26.09.2021). – Текст: электронный.

2. Chicago Manual of Style. – 17th e.d. – Chicago: University of Chicago Press, 2017. – 1146 p.

3. Harvard System of Referencing Guide. - 6th e.d. – Leabharlann UCD, 2018. – 24 p.

4. ГОСТ 7.1-2003. Система стандартов по информации, библиотечному и издательскому делу. Библиографическая запись. Библиографическое описание. Общие требования и правила составления. Принят Межгосударственным советом по стандартизации, метрологии и сертификации, протокол от 02.07.2003 №12-ст – URL: <http://docs.cntd.ru/document/1200034383> (дата обращения: 26.09.2021). – Текст: электронный.

5. ГОСТ Р 7.0.11-2011. Система стандартов по информации, библиотечному и издательскому делу. Диссертация и автореферат диссертации. Структура и правила оформления. Утверждён и введён в действие Приказом Федерального агентства по техническому регулированию и метрологии от 13 декабря 2011 г. N 811-ст – URL: <http://docs.cntd.ru/document/1200093432> (дата обращения: 26.01.2021). – Текст: электронный.