

Проблемы создания базы данных словаря сокращений

Acquiring the abbreviation dictionary database

Ю. В. Смирнов

*Государственная публичная научно-техническая библиотека России,
Москва, Россия*

Yury Smirnov

*Russian National Public Library for Science and Technology,
Moscow, Russia*

В докладе описывается база данных словаря сокращений, разрабатываемая в ГПНТБ России. Автор обосновывает необходимость создания такого словаря, а также приводит основные задачи, которые он должен будет решать. Основными источниками являются 2 государственных стандарта на сокращения слов и словосочетаний на русском и европейских языках, а также стандарт на единицы величин, не входящий в серию СИБИБД. Далее приводится набор полей базы данных, который был выработан на основе полученного материала, и статистика словаря на сегодняшний день. После описания основной информации о базе данных, автор делится основными проблемами, с которыми столкнулся во время работы над созданием словаря. В заключении приводятся основные задачи, которые еще следует решить в ближайшем будущем.

Ключевые слова: словари, сокращения, словари сокращений, базы данных, государственные стандарты.

The database of abbreviation dictionary being designed at the RNPLS&T is described. The relevancy of the dictionary is substantiated, and its main tasks are defined. The key sources for the dictionary are 2 federal standards of word and word combinations in Russian and European languages, and the non-SIBID standard for measurement units. The fields list based on these sources is given, and the dictionary statistics is cited. After the general information on the database, significant problems of acquiring the dictionary are described. The main tasks for the future are discussed.

Keywords: Dictionaries, abbreviations, dictionary of abbreviations, databases, national standards.

При работе (составление и идентификация издания) с библиографическими записями и библиографическими ссылками довольно часто используются сокращения. В качестве помощи работающим с ними людям в ГПНТБ России начал создаваться словарь сокращений слов и словосочетаний, который призван помочь пользователям решать следующие задачи:

- расшифровка вызывающих затруднения сокращений;
- поиск утвержденных государственными стандартами сокращений;
- поиск соответствий сокращений в русском и европейских языках;
- перевод сокращений на европейских языках на русский язык;
- создание новых сокращений, отсутствующих в государственных стандартах.

При работе с библиографическими записями и ссылками может использоваться не только русский язык, но и другие языки, поэтому словарь создается многоязычным.

База данных (БД) словаря создается с учетом действующих государственных стандартов, основными из которых являются:

- 1) ГОСТ Р 7.0.12-2011. СИБИБД. Библиографическая запись. Сокращение слов и словосочетаний на русском языке. Общие требования и правила [1];
- 2) ГОСТ 7.11-2004 (ИСО 832:1994). СИБИБД. Библиографическая запись. Сокращение слов и словосочетаний на иностранных европейских языках [2];
- 3) ГОСТ 8.417-2002. Государственная система обеспечения единства измерений. Единицы величин [3] (этот стандарт, хоть и не входит в систему СИБИБД, но включает множество сокращений, которые иногда могут вызывать сомнения в правильности их написания).

Первые два стандарта содержат приложения со списком слов и словосочетаний в полной и сокращенной формах, а также условия применения (например: «адмирал – адм. – при именах собст-

венных» и «административный – адм. – сокращается также в заголовке»). Для второго стандарта указан также код языка.

В последнем стандарте отсутствует такое приложение, поэтому для составления списка слов и их сокращений анализировался текст и различные таблицы стандарта.

В результате проделанной работы была разработана структура БД словаря сокращений, которая содержит следующие поля:

- 1) **Слово (словосочетание)** – содержит слова и словосочетания в полной форме на всех предлагаемых в словаре языках;
- 2) **Синоним** – представляет собой ссылку на полную форму слова или словосочетания;
- 3) **Сокращение** – содержит сокращение слова или словосочетания согласно стандартам;
- 4) **Код языка** – представляет собой трехбуквенный код языка по ГОСТ 7.75-97 [4];
- 5) **Условие применения** – содержит условия применения указанного сокращения;
- 6) **Примечание** – содержит примечания и пояснения к слову или словосочетанию, часто используется с единицами величин, поясняя тип единицы СИ, а также может содержать значение, соотношение с другими единицами. Это поле также используется для корректного отображения на экране надстрочных и подстрочных знаков для сокращений (например: «м²»);
- 7) **Перевод на русский язык** – используются для слов и словосочетаний на иностранных языках, и содержит ссылки на полную форму слова или словосочетания на русском языке;
- 8) **Источник** – Номер стандарта, откуда было взято слово или сокращение;
- 9) **Раздел** – раздел источника, из которого было взято слово или сокращение;
- 10) **Аналог** – ссылка на полную форму слова или словосочетания, по аналогии с которым было сделано сокращение.

На данный момент БД словаря содержит 8603 записи, которые можно разделить на несколько групп:

- из ГОСТ 7.11-2004 (ИСО 832:1994):
 - слова и словосочетания на языках, использующих латинский алфавит (21 язык): 6067 записей;
 - слова и словосочетания на языках, использующих кириллический алфавит (3 языка, кроме русского): 899 записей;
 - слова и словосочетания на греческом языке: 190 записей;
- слова и словосочетания на русском языке:
 - из Приложения к ГОСТ Р 7.0.12-2011: 618 записей;
 - из текста ГОСТ Р 7.0.12-2011: 24 записи;
 - из текста ГОСТ Р 7.0.100-2018 [5]: 2 записи;
 - слова и словосочетания, выделенные при переводе с европейских языков из ГОСТ 7.11-2004 (ИСО 832:1994):
 - сокращенные по правилам ГОСТ Р 7.0.12-2011: 262 записи;
 - сокращенные по аналогии со словами и словосочетаниями из ГОСТ Р 7.0.12-2011: 126 записи;
 - не сокращенные: 121 запись;
- из текста ГОСТ 8.417-2002 было выделено 122 записи на русском языке, которые были переведены на:
 - на английский язык: 70 записей;
 - на французский язык: 55 записей;
 - на немецкий язык: 53 записи;
- общепринятые и часто используемые сокращения на русском языке, но не вошедшие в стандарт (например: «и т.д.»): 7 записей.

подавляющее большинство записей не вызвало проблем при составлении базы данных словаря сокращений, поскольку в стандартах по сокращениям на русском [1] и европейских языках [2] содержатся хорошо проработанные таблицы, однако при переводе и дальнейшей работе с этими и другими стандартами пришлось столкнуться с некоторыми проблемами:

1. Поскольку ГОСТ 7.11-2004 (ИСО 832:1994) был создан на основе международного стандарта от 1994 г., то в него попали исторически устаревшие сокращения государств, которые отсутствуют в стандарте на сокращения на русском языке [1] (например: «Deutsche Demokratische Republic – Германская Демократическая Республика»). Было решено сохранить такого рода устаревшие слова, но снабдить их пометкой «Устаревшее».
2. В стандарте на сокращения на европейских языках [2] приводятся сокращения часто используемых слов с словосочетаний: «и другие» и «и так далее», однако в стандарте на русские сокращения [1] они отсутствуют. В этом случае был использован стандарт на библиографическое описание [5], где эти сокращения присутствуют.
3. Также во время работы были добавлены сокращения, отсутствующие в стандартах, но часто использующиеся (например: «до н.э.», «ср.»). В примечаниях к ним были добавлены соответствующие пометки.
4. Как уже говорилось выше, в стандарте на единицы величин [3] отсутствует единая таблица сокращений, поэтому выделение слов и их сокращений производилось по тексту и таблицам этого стандарта. Особенностью этого стандарта является наличие двух сокращений (русского и международного) для каждого термина на русском языке. Решение этого вопроса заключалось в присвоении двух сокращений для каждой единицы СИ в соответствии со стандартом, а также в переводе на английский, немецкий и французский языки, который был выполнен не полностью, а только для наиболее часто используемых единиц СИ. Причина этого заключается в отсутствии в международной системе стандарта полностью аналогичного русскому.
5. Перевод слов и сокращений на европейских языках на русский язык довольно часто был проблематичным, поскольку набор, подвергающихся сокращению слов на русском и европейском языках, неодинаков. Можно выделить три основных проблемных случая:
 - 1) В наборе русских слов и словосочетаний отсутствуют аналоги на европейских языках, поэтому эти переводы были добавлены в словарь и при возможности сокращены. Таких групп получилось три:
 - а) сокращены по правилам ГОСТ Р 7.0.12-2011 (например: «австралийский – австрал.»);
 - б) сокращены по аналогии с другими словами и словосочетаниями из ГОСТ Р 7.0.12-2011 (например: «автобиография – автобиогр. – по аналогии с ‘биография’»);
 - в) не сокращены (например: «июль», «иврит», «коллектив»);
 - 2) Крайне редко, но все же встречались слова, перевод которых не был осуществлен, поскольку в словарях отсутствовали эти слова (например: финское слово «toimittajakunta» или греческое – «σεδῖς»).
 - 3) Иногда вместо перевода с европейского языка был использован не прямой перевод, а ближайшее соответствие, которое выбиралось из списка слов и словосочетаний, существующих в ГОСТ Р 7.0.12-2011.

Несмотря на описанные выше сложности, база данных словаря сокращения была наполнена не только словами и словосочетаниями из стандартов, но и словами, которые появились при переводе стандарта на сокращения на европейских языках [2]. За счет этих слов база данных пополнилась на 566 слов и словосочетаний, имеющих сокращения.

В ближайшем будущем планируется дополнить эту базу данных сокращениями из действующих стандартов СИБИБД (добавится около 70 записей), а также сделать программный интерфейс для работы с этим словарем.

Список источников

1. **ГОСТ Р 7.0.12-2011.** СИБИД. Библиографическая запись. Сокращение слов и словосочетаний на русском языке. Общие требования и правила. – Введ. 2012-09-01. – М. : Стандартиформ, 2012. – 24 с.
2. **ГОСТ 7.11-2004 (ИСО 832:1994).** СИБИД. Библиографическая запись. Сокращение слов и словосочетаний на иностранных европейских языках. – Взамен ГОСТ 7.11-2004; Введ. 2005-09-01. – М. : Стандартиформ, 2005. – 82 с.
3. **ГОСТ 8.417-2002.** Государственная система обеспечения единства измерений. Единицы величин. – Введ. 2003-09-01. – М. : Стандартиформ, 2018. – 33 с.
4. **ГОСТ 7.75-97.** СИБИД. Коды наименований языков. – Введ. 1999-01-01. – М. : Изд-во стандартов, 1998. – 24 с.
5. **ГОСТ Р 7.0.100-2018.** СИБИД. Библиографическая запись. Библиографическое описание. Общие требования и правила составления. – Введ. 2019-07-01. – Москва : Стандартиформ, 2018. – 66 с.