

**Методы формирования многоязычного нормативного/
авторитетного файла географических названий**

**Methods of Formation of the Multilingual Normative/
Authority File of Geographic Names**

**Методи формування багатомовного нормативного/
авторитетного файлу географічних назв**

О. А. Лавренова

Российская государственная библиотека, Россия, Москва

Olga Lavrenova

Russian State Library, Moscow, Russia

О. О. Лавренкова

Російська державна бібліотека, Москва, Росія

Дается описание проекта тезауруса географических названий, формируемого в качестве основы национального авторитетного/нормативного файла. Тезаурус географических названий создается в интересах пользователей электронных каталогов, электронных библиотек и других автоматизированных информационных систем (АИС). Он предназначен для поддержки многоязычного доступа к данным с использованием контролируемых точек доступа, представляющих собой названия географических объектов.

Можно выделить следующие наиболее существенные характеристики данного нормативного файла: надежность источников информации; экономичность процессов формирования словаря, реализуемых преимущественно программно; высокие темпы пополнения базы данных и высокая полнота сведений о географических объектах России; учет смысловых связей между географическими названиями; многоязычность; открытый удаленный доступ. Нормативный файл установлен в открытом доступе на Web-сайте РГБ как одна из нормативно-справочных баз данных.

The project of the thesaurus of geographic names to make the foundation of the national authority/normative file is described. Users of electronic catalogs, e-libraries and other AIS will benefit from such thesaurus. It is designed to support multilingual access to the data using target access points that are geographic names.

The most important features of this normative file are the following: reliability of information sources; cost efficiency of the thesaurus acquisition accomplished mainly with software tools; high rate of database development and exhaustiveness of data on Russian geographic objects; control of relationship between geographic names; multilingual character; open remote access. The normative file is publicly accessible at the RLS' www-site as one of its normative reference databases.

Подано опис проекту тезауруса географічних назв, який формується як основа національного авторитетного/ нормативного файлу. Тезаурус географічних назв створюється для користувачів електронних каталогів, електронних бібліотек та інших автоматизованих інформаційних систем (АІС). Він призначений для підтримки багатомовного доступу до даних із використанням контрольованих пунктів доступу, які є назвами географічних об'єктів.

Можна виділити такі найсуттєвіші характеристики даного нормативного файлу: надійність джерел інформації; економічність процесів формування словника, які реалізуються переважно програмними засобами; високі темпи поповнення бази даних та високий ступінь відомостей про географічні об'єкти Росії; облік смислових зв'язків між географічними назвами; багатомовність; відкритий віддалений доступ. Нормативний файл встановлено у відкритому доступі на Web-сайті РДБ як одна із нормативно-довідкових баз даних.

1. Унификация элементов данных в библиографических записях (БЗ) – одно из главных направлений развития электронных каталогов. Использование в качестве содержимого полей БЗ некоторых принятых, санкционированных или стандартизированных элементов данных позволяет существенно повысить качественные показатели информационного поиска. Рассмотрим средства унификации одного из разновидностей элементов данных БЗ, а именно: географических названий.

В таблице 1 представлены типы обычно контролируемых элементов данных БЗ, которые могут содержать географические названия.

Таблица 1. Контроль географических названий в БЗ по авторитетным/ нормативным файлам

Элементы данных в полях БЗ	Авторитетные/нормативные файлы
Географическое название в составе наименования организации	Файл наименований организаций (начальный элемент ввода или географический подзаголовок)
Унифицированное заглавие	Файл унифицированных заглавий (географический подзаголовок унифицированного заглавия)
Географический заголовок картографического материала (заголовок описания картографического материала – структурированное географическое или тематическое наименование)	Файл заголовков картографических материалов (начальный элемент ввода или географический подзаголовок)
Имя лица	Файл имен лиц (географический подзаголовок)
Географическое название как предмет (тема)	Различные файлы предметных рубрик и другие тезаурусы

Принятая форма географического названия в качестве темы или предмета определяется используемой системой предметизации или тезаурусом, и поэтому ключевым вопросом качества используемых в БЗ элементов данных такого рода является **выбор источника данных для построения тезауруса. Одно дело, если составители** предметных рубрик выверяют каждое географическое название самостоятельно, пользуясь доступными источниками, другое дело, когда получают данные из официального источника, при создании которого изначально использовались достаточно редкие официальные документы и справочники.

2. Федеральный закон Российской Федерации «О наименованиях географических объектов» устанавливает правовые основы деятельности в области присвоения наименований географическим объектам и переименования географических объектов, а также нормализации, употребления, регистрации, учета и сохранения наименований географических объектов как составной части исторического и культурного наследия народов Российской Федерации. Федеральный закон ([4], ст. 8, п. 2) регламентирует следующее: «В документах, картографических, иных изданиях на русском языке или на других языках народов Российской Федерации употребляются нормализованные наименования географических объектов».

В связи с этим в РГБ принято следующее **основное требование к нормативным записям для названий российских географических объектов**: нормализованные (принятые) заголовки нормативных записей должны соответствовать нормализованным заголовкам, установленным на федеральном уровне и приведенным в автоматизированном **Государственном каталоге географических названий России (АГКГН)** [1-3], который формируется Государственным научно-исследовательским институтом геодезии, аэросъемки и картографии (ЦНИИГАиК). Тезаурус географических названий установлен в открытом доступе на Web-сайте РГБ в составе системы АЛЕФ, на которой основана библиотечная система РГБ, как общедоступный электронный ресурс [5]. Файл постоянно пополняется. На 1 мая 2008 г. получены и обработаны данные о 166000 географических объектов (70 регионов РФ). Российский нормативный файл географических названий получил международный код *rugeo* в системе форматов MARC21.

Пример сокращенной нормативной записи в формате MARC21 из файла РГБ:

```

151 \aБалахановское, озеро (находится на р. Вуокса) – принятый заголовок
451 \wa \aБалахановское (Торхон-ярви) – непринятые формы
451 \wa \aТарханьярви
451 \wa \aТорхон-Ярви
451 \wa \aТорхонь-Яр
451 \wa \aТурхонь Ярви
551 \wg \aЛенинградская, область (Россия). Природные объекты
670 \aСписки Ленинградского облисполкома, 1950 – источники данных
670 \aРесурсы поверхностных вод СССР, 1965

```

670 \aАтлас Российской Империи, 1804

670 \aАдминистративная карта Ленинградской области, 1986

670 \aОтвет Приозерского горисполкома от 26/03/1991.

670 \aАГКГН. ЦНИИГАиК: \b(Балахановское (Торхон-ярви) (31. 12. 1965). Тарханьярви (31. 12. 1945). Торхон-Ярви (31. 12. 1939). Торхонь-Яр (31. 12. 1899). Турхонь Ярви (31. 12. 1804). Балахановское (31. 12. 1950))

680 \i60° 47' СШ, 29° 45' ВД

Состав элементов данных в записях такого рода позволяет успешно выполнять **основные задачи авторитетных/нормативных файлов, определенные в модели FRAD** (Функциональные требования к нормативным данным): find (найти), identify (идентифицировать), contextualize (контекстуализировать), justify (обосновать, подтвердить).

3. Далее приведем **примеры возможностей использования записей** из файла при формировании библиографических записей.

Пример сокращенной библиографической записи на книгу в формате MARC21:

084|a ПЗ81. 1е(2)лб4я431(2) |2 rubbk – индекс полного варианта таблиц ББК

24500|a *Столетие опытных работ в Бузулукском бору*: |b Сб. ст. по науч. -практ. конф. |c [Редкол.: Родин С. А. и др.]

650 7|a *Сельское и лесное хозяйство. Сельскохозяйственные и лесохозяйственные науки -- Лесное хозяйство. Лесохозяйственные науки -- Отдельные группы и породы лесных деревьев и кустарников -- Хвойные породы -- Сосны -- Лесные заповедники. Лесопарки. Дендрарии -- Материалы конференции* |2 rubbk – иерархическая цепочка словесных формулировок индекса ББК

653|a *Бузулукский бор* – ключевое слово651 7|a *Бузулукский, район (Россия, Оренбургская область)* |2 rugeo – принятый заголовок из нормативного файла (тезауруса) географических названий.

Используемая нормативная запись в формате MARC21 (вывод на экран в стандартной форме системы АЛЕФ при поиске по признаку «Бузулукский»):

Заголовок: *Бузулукский, район (Россия, Оренбургская область)*

Примечание для пользователя: *район областное подчинение центр – город Бузулук*

Шире: *Оренбургская область (Россия)*

Пример сокращенной библиографической записи на нотное издание в формате MARC21:

084|a Ш92(2P=Рус.)-411я43 |2 rubbk

084|a Ш9(2P=Рус)-49 *Терский р-он, Мурманск. обл.* |2 rubbk

1001|a *Балашов, Дмитрий Михайлович*

24500|a *Русские свадебные песни Терского берега Белого моря* |b *Для пения (соло, ансамбль, хор) без сопровожд.* 651 7|a *Терский, берег (берег северной части Белого моря от м. Святой Нос до м. Лудошный)* |2 rugeo

651 7|a *Терский, район (Россия, Мурманская область)* |2 rugeo

651 7|a *Белое, море (внутреннее море Северного Ледовитого океана, между полуостровами Кольский и Канин)* |2 rugeo

Используемые нормативные записи в форме вывода системы АЛЕФ:

(а)

Заголовок: *Терский, берег (берег северной части Белого моря от м. Святой Нос до м. Лудошный)*

Неприн. форма: *Терский Берег*

ИсторСпра: Данные в АГКГН: *Терский Берег (31. 12. 1804). Терский берег (31. 12. 1939)* Прим. польз: 66 град. 20 мин. С. Ш., 40 град. 10 мин. В. Д.

Шире: *Мурманская область (Россия). Природные объекты*

(б)

Заголовок: *Белое, море (внутреннее море Северного Ледовитого океана, между полуостровами Кольский и Канин)*

Прим. польз: 66 град. 0 мин. С. Ш., 39 град. 0 мин. В. Д.

Шире: *Мурманская область (Россия). Природные объекты*

Шире: *Карелия, республика (Россия). Природные объекты*

Шире: *Архангельская, область (Россия). Природные объекты*

Шире: *Ненецкий, автономный округ (Россия). Природные объекты*

(в)

Заголовок: Терский, район (Россия, Мурманская область)

Прим. польз: район, областное подчинение, центр – рабочий поселок Умба

Шире: Мурманская область (Россия)

4. Для идентификации географического названия нередко необходимо уточнить объект, обозначаемый данным именем, поскольку в мире имеется большое количество совпадающих названий. Для этого служат пояснения в скобках, которые обычно используются в тезаурусах, а также примечания, имеющиеся в нормативной записи. Важным представляется одно обстоятельство, которое нередко игнорируется, в частности, в файлах предметных рубрик у нас и за рубежом: недопустимо утверждать, что нет необходимости уточнять значение географического названия, которое «всем известно», ибо не существует человека, который знает все географические названия даже родной страны, не говоря о других государствах. Например, считается, что озеро Байкал всемирно известно, но мало кто из россиян знает, что в Сибири существует несколько озер с этим названием. В результате подобных умозаключений часто бывает чрезвычайно трудно работать с зарубежными файлами, так как их составители, похоже, также пребывают в полной уверенности, что их города и реки должны знать все, а пояснения дают для зарубежных названий. В эпоху всемирного сетевого общения такие легкомысленные подходы неуместны.

Далее приведен пример уточнения географического названия «Фролищи» путем поиска в файле географических названий РГБ.

Вывод результата поиска на экран АЛЕФ:

- 1 Фролищи, поселок городского типа (Россия, Нижегородская область, Володарский район)
- 2 Флорищи, село (Россия, Владимирская область, Кольчугинский округ)

Вывод найденных нормативных записей на экран в стандартной форме системы АЛЕФ:

Заголовок	Фролищи , поселок городского типа (Россия, Нижегородская область, Володарский район)
Неприн. фор	Фролищева Пустынь, поселок городского типа Фролищева пустынь, поселок городского типа Фролищи (Фролищева Пустынь), поселок городского типа Флорищевая Пустынь, поселок городского типа Флорищи, поселок городского типа Данные в АГКГН: Фролищева Пустынь (31. 12. 1804). Фролищева пустынь (31. 12. 1941).
ИсторСпра	Фролищи (Фролищева Пустынь) (31. 12. 1957). Флорищевая Пустынь (31. 12. 1863). Флорищи (31. 12. 1939). Фролищи (15. 9. 1957)
Прим. польз	56° 25' СШ, 42° 40' ВД
Шире	Володарский, район (Россия, Нижегородская область)
Сист. номер	000003780
Заголовок	Флорищи, село (Россия, Владимирская область, Кольчугинский округ)
Неприн. фор	Фролищи , село
ИсторСпра	Данные в АГКГН: Фролищи (31. 12. 1863). Флорищи (1. 10. 1948)
Прим. польз	56 ° 18 ' С. Ш., 39 ° 14 ' В. Д.
Шире	Кольчугинский, округ (Россия, Владимирская область)

Таким образом, выяснилось, что для одного географического объекта, находящегося в Володарском районе Нижегородской области, форма «**Фролищи, поселок городского типа**» (с пояснениями) является принятой. Оказалось, что принятой формой названия другого географического объекта, находящегося в Кольчугинском округе Владимирской области, является название «**Флорищи, село**» (с пояснениями). Принятая форма была найдена по неприянтому варианту. «**Фролищи, село**» Для копирования конкретной принятой формы из нормативной записи в библиографическую остается сделать выбор на основании содержания каталогизируемого документа.

5. Проект данного нормативного файла РГБ ведет с 2003 года. До 2007 года тезаурус формировался только для географических названий России по технологии, опирающейся, в первую очередь, на использование данных из автоматизированного Государственного каталога географиче

ческих названий. При этом в 2006 г. начат эксперимент по дополнению тезауруса названиями на государственных языках субъектов РФ в оригинальной графике [1-3]. **Данные о географических названиях на государственных языках субъектов РФ**, отличных от русского, в оригинальной графике будут, надо полагать, предоставлять национальные библиотеки. Первый эксперимент проводится совместно с Национальной библиотекой Чувашии.

Данные о зарубежных географических названиях планируется получать следующими способами: из зарубежных национальных библиотек (передача файлов, обмен данными); путем участия в международном проекте VIAF (в будущем); из отечественных библиотек – в рамках технологического процесса Сводного каталога библиотек России – СКБР (технология проектируется); путем интерактивного формирования записей в РГБ.

Эксперимент по работе с зарубежными нормативными файлами географических названий проводится в порядке научного сотрудничества с Библиотекой Конгресса США (БК), из которой получено около 150000 нормативных записей. Специалисты РГБ с использованием специально разработанного программного обеспечения производят сравнение нормативных записей на российские географические названия из файлов БК и РГБ, устанавливая связи между записями для одних и тех же географических объектов. В поле связи записи из одного файла вносятся контрольный номер и форма принятого заголовка из другого файла. Что касается названий зарубежных географических объектов, то имеющиеся для них записи в файле БК преобразуются (частично программно) в записи файла РГБ, если это заимствование признается целесообразным. Естественно, заново формируются «формальные» поля данных. Заголовок переводится на русский язык с использованием картографических материалов, справочных изданий и т. д. Переводу подлежат также определенные типы примечаний. Принятый заголовок записи БК становится непринятой формой в записи РГБ. Кроме того, сохраняются непринятые формы из записи БК и указания на источники данных в оригинале. Программно вносится поле источника «*Library of Congress Authorities*» (*авторитетные/нормативные файлы Библиотеки Конгресса*). Создается поле более широкого понятия по правилам, установленным для файла РГБ. Программно удаляются ненужные поля и устанавливаются связи между принятыми заголовками РГБ и БК. Обе библиотеки работают с форматом MARC21, поэтому чужие записи легко воспринимаются партнером и его программными продуктами.

Пример преобразованной записи БК с принятым заголовком «*Hansaviertel (Berlin)*» в сокращенно форме:

#151 \aХансафиртель, район (Германия, Берлин)

#4101 \aHansaviertel (Berlin)

#551 \aБерлин, город (Германия)

#670 \a Library of Congress Authorities

#670 \aDolff-Bonekampfer, Gabriele. Das Hansaviertel, c1999: \bрайон Берлина, сильно разрушенный во время войны, восстановленный, и сейчас в нем размещены 343 здания.

#670 \aHansaviertel WWW site, Sept. 7, 2000 \bЗдания построены как часть программы реставрации выставки Интербау, в которой приняли участие многие иностранные архитекторы.

#751_0 \aHansaviertel (Berlin) \bn 00006757 – связь с заголовком в файле БК.

6. **Дальнейшая работа с файлом** заключается преимущественно в следующем:

- выявление в записях непринятых российских названий, которые в БЗ каталогов РГБ ранее признавались принятыми и для которых следует создать собственные нормативные записи (например, *Ленинград*, *Свердловск*);

- включение в записи вариантов названий объектов РФ на других языках в оригинальной графике;

- ввод в файл названий зарубежных географических объектов на русском и иностранных языках;

- дополнение записей прежними географическими названиями, если они отсутствовали в исходных записях;

- формирование нормативных записей на исторические названия исчезнувших географических объектов;

- работы по внедрению национального нормативного файла географических названий в СКБР.

7. Проект финансируется:

- из бюджета РГБ,
- по договорам с Министерством культуры РФ в 2003 г. и с ФАКК в 2004 г.,
- по государственным контрактам с ФАКК в 2006, 2007 и 2008 г. г.

Автор благодарит сотрудников РГБ Аветисову Т. В., Котельникову Н. Е. и Зинчук Л. Н., а также Центральный научно-исследовательский институт геодезии, аэросъемки и картографии, без которых успешная реализация проекта не представляется возможной.

Литература

1. Лавренова О. А. Многоязычный доступ к данным на основе тезауруса географических названий // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. Труды Девятой Всероссийской научной конференции (RCDL'2007). Переславль-Залесский, 15-18 октября 2007 г. Т. 2 – Переславль-Залесский: 2007. – С. 57-62
2. Лавренова О. А. Национальный нормативный файл географических названий России // Международный библиотечный и информационный конгресс: 71-я генеральная конференция ИФЛА. 14-18 августа 2005. Осло, Норвегия. – Материалы. – Русская версия: http://www.ifla.org/IV/ifla71/papers/015r_trans-Lavrenova.pdf. Английская версия: <http://www.ifla.org/IV/ifla71/papers/015e-Lavrenova.pdf>
3. Лавренова О. А. Национальный файл географических названий – новый проект РГБ // Библиотекосведение. – 2006. – №2. – С. 46-53
4. Российская Федерация. Федеральный закон от 18 декабря 1997 г. № 152-ФЗ. «О наименованиях географических объектов»
5. Файл нормативных/авторитетных записей географических названий России в электронном каталоге РГБ: <http://aleph.rsl.ru/> (БАЗЫ ДАННЫХ/СПРАВОЧНИКИ/ Географические названия).