

ЛИНГВИСТИЧЕСКАЯ МОДЕЛЬ АВТОМАТИЧЕСКОГО СЛОВАРЯ ТОПОНИМОВ АРИЗОНЫ

На сегодняшний день создание электронных информационных систем является наиболее продуктивным способом предоставления доступа к региональному топонимическому материалу и его статистической обработки и интерпретации. Принципиальное разнообразие лингвистической и экстралингвистической информации, закодированной в именах географических объектов, а также системный характер топонимии в целом требуют обращения современных топонимистов к способам ее структурирования в контексте задач компьютерной лексикографии, содействующей обработке и систематизации больших массивов данных, хранимых и управляемых с помощью компьютера.

Опишем кратко особенности построения словарной статьи и принципы организации данных в моделируемом словаре топонимов штата Аризона (США), а также представим основные аспекты компьютерной обработки данных топонимии с помощью созданной нами базы данных. Материалом для создания словаря послужили собственные имена региона, извлеченные из словаря Б. Х. Грэнджера «Arizona's Names».

Предварительная подготовка к составлению электронного словаря топонимов предполагает принятие решений исследователя по трем аспектам: 1) определение объектов описания; 2) выбор системы управления базой данных (далее – СУБД) топонимии; 3) организация макро- и микроструктуры моделируемого словаря. Рассмотрим их последовательно.

Анализ структурных и семантических особенностей материала исследования в нашей работе был ограничен собственными именами населенных пунктов (ойконимы), рек (потамонимы) и гор (оронимы). Согласно объекту лексикографического описания, данный электронный словарь является приватным словарем топонимов. Выборка названий осуществлялась в ходе работы со словарем и была механической или систематической, при которой первый элемент отбирался случайно, затем с шагом n отбирался каждый k

элемент. Размер генеральной совокупности при этом $N = nk$. В итоге реестр названий включил в себя каждое нечетное название со страницы. Объем полученной базы данных топонимов составил 649 единиц.

Выбор СУБД определялся простотой в использовании и эффективностью при работе с информацией определенного объема (до 5 Гб). Этим критериям удовлетворяет MySQL. Для разработки и администрирования базы данных MySQL нами использовалась графическая оболочка dbForge, благодаря которой автоматизировался рутинный процесс переноса информации из печатной формы в электронную, а также появилась возможность создавать SQL-запросы, не выходя из программы, и получать нужную информацию. Если пользователь не владеет MySQL, то может воспользоваться СУБД Microsoft Access, которая устанавливается с каждым пакетом Microsoft Office. Данная СУБД также позволяет осуществить запрос на выборку или воспользоваться функцией «Конструктор».

Топонимы Аризоны размещались в словаре в алфавитном порядке, что свидетельствует об алфавитном типе организации макроструктуры. Микроструктура состоит из нескольких зон. Первая – зона номинации, здесь представлено собственное имя географического объекта. Далее следует зона экстралингвистической информации, в которой представлены название округа, буквенно-цифровое обозначения населенного пункта, а также района. Данные зоны являются значимыми, так как это дает возможность анализировать ограниченную группу названий определенного округа или района. Следующая зона – зона пометок. В ней представлена этимологическая информация об имени, отсылающая к его происхождению, что является важным для определения мотивационной базы топонима. Четвертая зона представлена типом структуры названия. Здесь топонимы подразделены на простые, простые суффиксальные, составные и сложные. Пятая и последняя зона – зона дериватов, представленная видами мотивационных основ, заложенных в названии объекта. Обращение к ней позволяет составить выборку топонимов апеллятивного либо ономастического происхождения.

Таким образом, при составлении топонимического словаря нами были учтены такие принципы лексикографического описания, как номинативный, географический, этимологический и словообразовательный.

Логичной итерацией, или новым циклом развития созданного словаря, будет его перенос в форму веб- или мобильного приложения. Веб-приложение – это клиент-серверное приложение, где обмен данными происходит по сети, а хранение осуществляется на сервере. Для использования мобильного приложения Интернет не нужен, так как все данные переносятся в процессе установки на само устройство. Однако для создания одного из таких приложений требуется время и знание нескольких языков программирования. По этой причине доступ к конечному файлу и работа над ним осуществляется в офлайн-формате при поддержке сторонних программ – dbForge или Microsoft Access соответственно.

Особенности использования полученного словаря стоит рассмотреть на примерах использования двух программ, описанных выше.

Первый из примеров – выборка ста четных единиц с указанием исключительно их этимологии и мотивационных основ для исследования связей между языком и культурой региона в процессе номинации. Чтобы получить названия, стоящие под четным номером в таблице, нам нужно, чтобы целочисленный остаток от деления этого числа равнялся нулю. Вот какой код у нас получится:

```
USE arizona;  
SELECT  
`№`,  
Title,  
Etymology,  
`Motivational base`  
FROM toponyms  
WHERE (`№` % 2) = 0  
LIMIT 100;
```

Для выборки названий, стоящих под нечетным номером, нужно поменять строчку с оператором 'WHERE' на 'WHERE (`№` % 2) > 0'.

Так работают запросы с применением оболочки dbForge. В программе Microsoft Access процесс создания запросов также легко осуществим в режиме конструктора. Смоделируем ситуацию: для проекта по лингвострановедению второго иностранного языка (английский) нужно подготовить информацию о горах Аризоны (так как территория штата находится в регионе под названием Basin and Ringe) и сделать вывод, является ли количество гор большим для данного региона по сравнению с другими.

Код, соответственно, имеет такой вид (его можно посмотреть в режиме SQL):

```
SELECT [toponyms].Title, [toponyms].Etymology  
FROM [toponyms]  
WHERE ((([toponyms].Title) Like "* Mountain")) OR ((([toponyms].Title) Like "* Mountains")) OR ((([toponyms].Title) Like "Mount *"));
```

Как видно из примера выше, нами выбраны только 2 поля – «Название» и «Этимология», как того требует задание. Условия отбора включают в себя шаблон, в котором название будет включать слова *mountain/mountains/mount* 'гора, горная цепь, горы'.

Из сказанного выше следует, что, используя данный электронный словарь, можно выполнять различные запросы, касающиеся лингвистических и экстралингвистических характеристик топонимов Аризоны, и получать соответствующие результаты с возможностью последующего сохранения.

Электронный словарь обладает особенностями, выгодно отличающими его от печатного словаря, – такими как быстрый поиск, гипертекстовость, универсальность, мультимедийность, интерактивность и возможность быстрого редактирования содержимого. Обращение к электронной форме позволяет исследователю топонимии получить результат в считанные секунды (среднее время выполнения запроса – 0,007 секунд).

Создание автоматического словаря топонимов Аризоны решает такие задачи, как 1) фиксация единиц языкового уровня в удобном для современных исследований формате; 2) сохранение языковых данных; 3) оптимизация исследований для получения новых результатов; 4) верификация результатов исследования за счет обращения к материалу большого объема.

Представленный нами автоматический словарь топонимов Аризоны может использоваться как для анализа языковых и внеязыковых характеристик географических имен штата (семантика, структура и др.), так и в качестве модели для создания электронных словарей других регионов (в т. ч. Беларуси). Кроме того, результаты работы могут быть использованы на занятиях по лингвострановедению и компьютерной лингвистике, при проведении спецкурсов по топонимике (студенты в таком случае получают неограниченный доступ к электронному ресурсу, что способствует формированию целостного представления об изучаемом предмете).