

УДК 81

<https://doi.org/10.33619/2414-2948/52/77>

СОСТАВЛЕНИЕ СЛОВАРЯ: КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ, ЭТАПЫ, ФУНКЦИОНИРОВАНИЕ

©Аманалиева Г. Э., ORCID: 0000-0002-8564-3288, Кыргызско-российский славянский университет, г. Бишкек, Кыргызстан, gulnara040805@mail.ru

DRAFTING OF DICTIONARY: CONCEPTUAL MODEL, STAGES, FUNCTIONING

©Amanaliyeva G., ORCID: 0000-0002-8564-3288, Kyrgyz-Russian Slavic University, Bishkek, Kyrgyzstan, gulnara040805@mail.ru

Аннотация. В настоящей статье описаны этапы составления словаря: обработка источников информации для словаря; подготовка начальной картотеки; подготовка проекта словаря (инструкции для составителей); написание и редактирование словаря; набор, верстка и издание книги; внесение дополнений, исправление ошибок с целью подготовки следующих изданий. Словарь представлен как многоплановая самостоятельная система с многочисленными внутренними связями, микро– и макроуровнями.

Abstract. The stages of drafting of dictionary: treatment of information generators for a dictionary; preparation of initial card index; preparation of project of dictionary (instructions for compilers); writing and editing of dictionary; set, make-up and edition of book; bringing of additions, correction of errors with the purpose of preparation of next editions are described in the article. A dictionary is presented as a many planned independent system with numerous intracommunications, micro– and by macrolevels.

Ключевые слова: лексика, словарь, термины.

Keywords: vocabulary, dictionary, terms.

Составление словарей на основе словарных баз данных — это современный и универсальный подход компьютерной лексикографии, но он требует создания достаточно большого и сложного программного обеспечения. В принципе, можно рассматривать задачу разработки программного комплекса для автоматизации лексикографических работ во всех ее аспектах. Хотя такая технология и является теоретически возможной, но на практике — это трудно решаемая и нецелесообразная задача. Это обусловлено целым рядом факторов.

С точки зрения лексикографа составление словаря можно разбить на ряд следующих этапов [1]:

- обработка источников информации для словаря;
- подготовка начальной картотеки;
- подготовка проекта словаря (инструкции для составителей)
- написание и редактирование словаря;
- набор, верстка и издание книги;
- внесение дополнений, исправление ошибок с целью подготовки следующих изданий.

Каждый из этих этапов требует той или иной степени автоматизации, представляя самостоятельную, относительно независимую и достаточно сложную задачу, которая



является предметом разработки соответствующего коллектива высококвалифицированных специалистов. Поэтому для каждого этапа разработки требуется достаточно специфическое и сложное программное обеспечение.

Ряд функций, которые должны входить в программный комплекс автоматизации лексикографических работ, обеспечиваются существующими программными пакетами (например, полиграфическое оформление текста обеспечивается серийными издательскими системами). Дублирование таких функций является нецелесообразным, как с точки зрения качества исполнения и скорости реализации, так и с точки зрения полноты функционирования, тем более, что такие пакеты являются сами по себе сложными в использовании и требуют соответствующей квалификации у пользователя.

С точки зрения разработчика программного обеспечения словарь в его компьютерном виде — это база данных, поэтому создание программного обеспечения для автоматизации лексикографических работ полностью соответствует требованиям к созданию программного обеспечения информационных систем [2]. С этой точки зрения в лексикографической деятельности можно выделить следующие составляющие:

- построение информационной модели словаря;
- реализация информационной модели средствами одной из существующих баз данных;
- создание соответствующего программного обеспечения;
- наполнение и сопровождение базы данных словаря;
- подготовка к полиграфическому изданию словаря того или иного типа на основе информации из базы данных;
- построение на основе базы данных различных информационно-поисковых и информационно-справочных систем;
- использование базы данных словаря в лексикологических исследованиях;
- использование базы данных словаря как информационной базы в программном обеспечении другого назначения.

В отношении обеспечения эффективности функционирования каждая такая составляющая имеет свои требования по оптимальной структуре данных, объемам памяти, скорости реакции на запросы пользователя и тому подобное. Эти требования сильно отличаются для разных составляющих, а часто и несовместимы между собой. Если на этапе формирования и сопровождения баз данных важнейшим является обеспечение целостности базы данных и менее важны время реакции системы и объем памяти, которую занимает база данных, и практически нет требований к объему программного кода, то при функционировании словаря в виде информационной системы на первый план выходят время реакции и объемы памяти. Если же словарь будет функционировать как составная часть другой системы, то в этом случае будут прилагаться дополнительные требования к объему программного кода, обеспечивающего функционирование словаря.

Специфической особенностью словарных баз данных по отношению к другим является относительно редкое изменение информации после первоначального наполнения и узкий круг специалистов, которые эти изменения вносят (словарь является авторской работой, потому разрешение вносить в него изменения другим лицам, кроме авторов, выглядит проблематичным).

Приведенные причины делают комплекс программ сложным программным продуктом как с точки зрения использования, так и внутренней структуры. Конкретный пользователь для решения своей конкретной задачи использовал бы только необходимое ему подмножество возможностей комплекса, обязательно согласовывая свои действия с условиями, налагаемыми другими, вполне ему не нужными, возможностями этого комплекса.

Функционирование такого комплекса возможно лишь в специализированных организациях с соответствующим техническим обеспечением и соответствующими кадрами, что, в свою очередь, утрудняет доступ широкого круга пользователей к современным инструментальным программным средствам такого типа.

Поэтому целесообразно задачу автоматизации лексикографических работ разбить на ряд независимых подзадач. Для каждой подзадачи создается своя концептуальная модель базы данных и свой независимый комплекс программ, функции и структура данных которого лучше всего соответствуют подзадаче. Отдельные программные комплексы связываются между собой на уровне обмена данными: программное обеспечение каждой подзадачи должно позволять экспорт / импорт данных в соответствии с требованиями непосредственно связанных с ней подзадач. Такой подход позволит значительно упростить структуру программного обеспечения, уменьшить время его создания и увеличить надежность функционирования [3]. Пользователь сможет выбирать только те средства, которые непосредственно необходимы для его работы. В специализированных организациях такие программные комплексы можно объединять в один на уровне операционной системы компьютера [4, 5].

При таком подходе становится возможным для всех приложений словаря иметь единый начальный этап — этап создания и наполнения словарной базы данных, причем, данные для этого этапа могут поступать различными способами: как результат работы программ автоматизированного анализа текста, как результат анализа других словарей, в результате ручного ввода информации пользователем и тому подобное. На этом этапе создается упрощенная структура базы данных, что значительно облегчает работу составителей и уменьшает сложность программного обеспечения, необходимого для этого этапа. Все будущие информационные и текстологические изменения и дополнения вносятся исключительно в базы данных этого первичного этапа. Проводится комплексный анализ структурной и семантической корректности собранной информации. Для каждого конкретного применения создаются дополнительные специализированные программные средства, которые будут осуществлять реструктуризацию данных в структуру, оптимальную и удобную именно для этого приложения. Такой подход позволит гарантировать идентичность информации во всех последующих приложениях словаря.

Необходимость существования единой начальной словарной базы обусловлена еще и тем, что в традиционной лексикографии информация об одном и том же элементе реестра в разных типах словарей содержит различную, а порой и противоречивую информацию [6]. Причин такой неоднозначности несколько:

- различные типы словарей готовятся различными авторскими коллективами;
- авторские коллективы имеют разные взгляды на одни и те же филологические явления;
- авторские коллективы не всегда имеют возможность обмениваться оперативной информацией;
- объективные технические причины (например, без наличия единого абсолютно полного источника информации про лексикографические единицы, которые включаются в словарь, практически невозможно сделать их описание идентичным во всех словарях).

Образование различных типов словарей на основе единой начальной словарной базы данных позволяет минимизировать данную проблему.

Таким образом, автоматизацию лексикографических работ будем рассматривать как технологическую линию, где на разных этапах применяются различные пакеты программ, как специализированные, так и общие, комплексно внедряют в лексикографию новейшие

информационные технологии. Перечислим еще некоторые обязательные требования к такой технологической линии:

-взаимосвязь между составляющими такой линии происходит на уровне обмена данными;

-пакеты программ должны быть удобным в использовании и отвечать современным требованиям «дружественности» интерфейса;

-пакеты программ должны позволять безболезненное наращивание возможностей.

Выделим следующие составные части технологической линии составления словарей на компьютере:

Подготовительная работа. Этому этапу соответствуют программные средства для автоматизации обработки источников к составлению словаря и созданию начальной картотеки.

Наполнение и сопровождение исходной базы данных. Программные средства этого этапа должны позволять ручной ввод информации, импорт данных из предыдущего этапа, редактирование и дополнение введенной информации, комплексную проверку структурной и семантической корректности, непротиворечивости и полноты собранных данных.

Реструктуризация лексикографической базы данных и выбор данных для конкретных приложений. На этом этапе программные средства должны с лексикографической базы данных создавать информационные базы конкретных приложений (например, текст словаря для полиграфии или справочная словарная система).

Конкретные применения собранных данных. Программные средства этого этапа позволяют использовать реструктуризированные данные конкретным образом:

-для изготовления оригинал-макета полиграфического издания словаря определенного типа,

-как специализированную информационно-справочную систему;

-как составную часть других программных комплексов;

-как входные данные различных лингвистических исследований и тому подобное.

Центральной составляющей технологической линии является словарная база данных, которую назовем лексикографической базой данных (ЛБД). Отметим, что с точки зрения информационного наполнения ЛБД не рассматривается как та, что отвечает одному конкретному словарю. Лексикографическая база данных содержит широкую совокупность лексикографической информации, включая рабочую, поэтому на основе нее можно получить целую совокупность словарей. Таким образом, основной задачей создания технологической линии является реализация ЛБД на компьютере.

Построение концептуальной информационной модели словаря. Информационную модель словаря будем строить исходя из предположения, что всякий словарь является определенной подмножеством универсального словаря. Основные структуротворческие элементы будем выделять на основе анализа структуры существующих словарей.

Объекты словаря и отношение между ними. Для построения концептуальной информационной модели словаря, в первую очередь, необходимо определить его объекты и отношения между ними.

Как было отмечено выше, словарь является многоплановой самостоятельной системой с многочисленными внутренними связями, в которой можно выделить микро- и макроуровни.

На макроуровне словарь будем рассматривать как объединение основы словаря и вспомогательных объектов. Основа словаря — это множество всех его статей. Вспомогательными объектами словаря источники словаря, предисловие к словарю, научный

аппарат словаря, описание его структуры, правила пользования словарем, список сокращений, таблицы, индексы, обратные указатели, ключевые слова и тому подобное. Одну часть этих объектов (например, разного рода индексы) можно получить на базе информации, содержащейся в основе словаря, другую – необходимо создать автономно (например, предисловие пишется непосредственно авторами словаря). Обязательной составляющей макроуровня словаря является его основа, поскольку без нее Словарь не будет словарем. Наличие всех остальных регламентируется такой характеристикой макроуровня, как концепция словаря — его назначение, круг пользователей, способ использования, ориентировочный объем, количество языков, входов и тому подобное. Поэтому все отношения этого уровня имеют единый характер – «присутствует в словаре».

Микроуровень непосредственно относится к основе словаря. Для определения объектов этих уровней и отношений между ними проанализируем структуру нескольких статей переводного и толкового словарей. Такой выбор обусловлен тем, что структура статей этих словарей практически отражает структуру словарей остальных типов.

При построении концептуальной информационной модели словаря необходимо учитывать то, что внутренняя структура ее объектов и способы фиксации взаимосвязей между ними должны позволять следующее:

-формирование на основе раз введенной информации как основной статьи, так и статьи-ссылки;

-включение при необходимости в реестр словаря отдельных лексикографических параметров (например, парадигму местоимений);

-образование при необходимости реестра какого-то лексикографического параметра (инверсия переводного словаря);

-иное использование отдельных элементов лексикографических параметров описательного типа (например, часть толкования в энциклопедическом и наоборот);

-реализовывать существующие словарные благоустройства.

Учитывая многоплановость структуры словаря, изобразим его концептуальную информационную модель в виде иерархии лексикографических систем. Обязательными являются первый уровень иерархии, где словарь представлен в виде элементарной лексикографической системы, которая соответствует его макроуровню, и второй – где основа словаря представлена как лексикографическая система, что соответствует микроуровню словаря. Низкие уровни иерархии соответствуют отдельным лексикографическим параметрам, представление которых в виде лексикографических систем зависит от степени детализации их отражение в ЛБД.

Таким образом, моделируя концептуальную модель, учитываем тот факт, что модель конкретного словаря не обязательно содержит все компоненты универсального словаря, поэтому в теоретической части диссертационного исследования будем говорить о базовых правилах построения концептуальной модели Учебного Словаря.

Список литературы:

1. Андриященко В. М. Концепция и архитектура машинного фонда русского языка. М.: Наука. 2009.
2. Дейт К. Дж. Введение в системы баз данных. М.: Вильямс. 2011.
3. Майерс Г. Надежность программного обеспечения. М.: Мир. 2018.
4. Гладков С. А., Фролов Г. В. Программирование в Microsoft Windows. М.: Диалог-МИФИ. 2012.

5. William H. Murray III and Chris M. Pappas. Windows programming: an introduction. USA: Osborne Mc. Graw-Hill, 2010.

6. Хойнацкий М. С. О правильности употребления и написания терминов в научно-технических, энциклопедических и орфографических словарях // Подготовка и использование научно-технических словарей в системе информационного обеспечения: всесоюзная конференция. М.: Русский язык. 2016.

References:

1. Andryushchenko, V. M. (2009). Kontseptsiya i arkhitektura mashinnogo fonda russkogo yazyka. Moscow.

2. Deit, K. Dzh. (2011). Vvedenie v sistemy baz dannykh. Moscow.

3. Maiers, G. (2018). Nadezhnost' programmnoho obespecheniya. Moscow.

4. Gladkov, S. A., & Frolov, G. V. (2012). Programmirovaniye v Microsoft Windows. Moscow.

5. William, H. (2010). Murray III and Chris M. Pappas. *Windows programming: an introduction. USA: Osborne Mc. Graw-Hill.*

6. Khoinatskii, M. S. (2016). O pravil'nosti upotrebleniya i napisaniya terminov v nauchno-tekhnicheskikh, entsiklopedicheskikh i orfograficheskikh slovaryakh. In *Podgotovka i ispol'zovanie nauchno-tekhnicheskikh slovarei v sisteme informatsionnogo obespecheniya: vsesoyuznaya konferentsiya, Moscow.*

*Работа поступила
в редакцию 05.02.2020 г.*

*Принята к публикации
12.02.2020 г.*

Ссылка для цитирования:

Аманалиева Г. Э. Составление словаря: концептуальная модель, этапы, функционирование // Бюллетень науки и практики. 2020. Т. 6. №3. С. 628-633. <https://doi.org/10.33619/2414-2948/52/77>

Cite as (APA):

Amanalieva, G. (2020). Drafting of Dictionary: Conceptual Model, Stages, Functioning. *Bulletin of Science and Practice*, 6(3), 628-633. <https://doi.org/10.33619/2414-2948/52/77> (in Russian).