

ИНФОЛОГИЧЕСКИЕ МОДЕЛИ ЯДЕРНЫХ ДАННЫХ ФОРМАТА ENDF

А.Г. Юферов

*Обнинский институт ядерной энергетики (ИАТЭ НИЯУ МИФИ)
249040, Калужская обл., г. Обнинск, Студгородок, 1*

Р

Рассмотрены вопросы инфологического моделирования библиотек ядерных данных формата ENDF с целью преобразования файлов ENDF в реляционную базу данных. Перенос в реляционный формат позволит использовать для обработки ядерных данных стандартные общедоступные средства, упрощающие освоение и эксплуатацию этого информационного массива. Инфологические модели описаны посредством формул вида «Сущность (список атрибутов)». Предлагаемые инфологические формулы построены исходя из физической природы данных и определяющих теоретических соотношений. Этим исключается введение специальной нотации для описания структуры и смысла данных. В результате упрощается применение реляционных форматов в проектных кодах и решение задач оценивания ядерных данных. Сформулирована концепция ядерной информатики на основе технологий реляционных СУБД как одного из инструментов решения проблемы «больших данных» в современной научно-технологической сфере. Указаны организационно-технологические основания для перевода библиотек ENDF в реляционный формат. Перечислены требования к форматам представления ядерных данных, обеспечиваемые реляционными СУБД. Отмечены особенности построения инфологических моделей, обусловленные иерархической природой ядерных данных. Указан порядок сохранения метаданных ENDF, что может быть полезно для верификации и валидации (проверки структурно-синтаксической правильности и работоспособности) как исходных данных, так и процедур конвертации в реляционный формат. Приведены формулы инфологических моделей для файла сечений, файла энергетических распределений вторичных нейтронов и файла энергоугловых распределений продуктов ядерных реакций. Полный массив инфологических моделей для библиотек ENDF и модули генерации соответствующих реляционных таблиц размещены на общедоступном сайте.

Ключевые слова: библиотеки ENDF, инфологическое моделирование, реляционный формат.

ВВЕДЕНИЕ

В работе продолжено рассмотрение вопросов, связанных с конвертацией файлов ядерных данных из библиотек ENDF в формат реляционных баз данных (РБД). Концепция и технология, а также программы и результаты конвертации библиотек РОСФОНД-А и ENDF/B-VII.1 были описаны в [1 – 3]. Полученный опыт указывает, что целесообразно сформулировать некоторые общие принципы описания ядерных данных, позволяющие стандартизировать и унифицировать перенос данных в РБД.

© А.Г. Юферов, 2018

Разработанные ранее алгоритмы и программы переноса использовали в качестве исходной модели формат ENDF как таковой. Однако для более эффективного применения реляционных форматов в проектных кодах, а также для проверки структурно-синтаксической правильности данных (верификации) и установления их адекватности решаемым задачами (валидации) целесообразно построить инфологические модели исходя из физической природы данных и определяющих соотношений. Под инфологической моделью здесь понимается описание данных в терминах объектов предметной области, их свойств и связей, ориентированное на последующее реляционное (табличное) представление и хранение данных.

КОНЦЕПЦИЯ ЯДЕРНОЙ ИНФОРМАТИКИ

Объем ядерных данных, разнообразие их структур сегодня столь велики, что обусловили формирование новой научной дисциплины – ядерной информатики. Цель этой дисциплины – создание компьютерных моделей ядерных данных исходя из их физического смысла и теоретического описания, а также из потребностей использования данных в качестве объекта анализа или массива констант для обеспечения проектных расчетов. Оба направления предполагают решение задач хранения, поиска, передачи, визуализации и обработки на основе современных информационных технологий. Тем самым достигается, в частности, новое качество в решении традиционных задач физика-оценщика, связанных с систематикой, верификацией и валидацией данных [4 – 6].

Проблема больших объемов научных данных возникла достаточно давно. Однако в силу исторических причин и обстоятельств организационного плана эта задача решается специфическим образом в каждой научно-технологической сфере. Так, формат GRIB был предложен для хранения метеоданных, формат netCDF – геоинформационных данных, формат FITS – астрономических и т.д. [7 – 9]. Идеология этих форматов аналогична стилю формата ENDF и предусматривает хранение в одном файле как массивов данных, так и соответствующей метаинформации в виде флагов, ключей, указателей, счетчиков, комментариев и т.д. Такой подход с необходимостью обуславливает создание специального языка и инструментария для каждого формата [10 – 13]. Однако интенсификация межотраслевого обмена данными, коммерциализация научных результатов требуют дальнейших усилий по унификации технологии работы с данными любой природы. Это особенно важно для ядерной энергетики, где физические данные представляют собой не только предмет научного анализа, но и константную основу проектных расчетов.

Реляционные СУБД могут быть основным инструментом ядерной информатики, поскольку вполне обеспечивают известные требования к форматам представления научных данных:

- возможность отображения данных разных типов и структур;
- поддержка больших объемов данных;
- мобильность данных, т.е. возможность обработки и хранения на различных программно-аппаратных платформах;
- быстрый ввод/вывод и визуализация;
- веб-доступ и веб-обработка;
- расширяемость формата.

Можно указать два основных «потребительских» аспекта технологии реляционных таблиц, обуславливающих эффективность работы с большими массивами ядерных данных:

- для физика-оценщика – это средство визуализации, полный аналог «бумажных» таблиц, обеспечивающий к тому же мгновенную выборку сечений таблицы по

строкам и столбцам, вывод их на график, оперативный расчет балансов и параллельный анализ различных версий данных;

– для физика-программиста – это возможность адресации и выборки данных «по имени» без контроля их фактического размещения во внешней или в оперативной памяти, что обеспечивает относительную независимость задач организации данных и их обработки.

ОСНОВАНИЯ И ЦЕЛИ ПЕРЕХОДА К РЕЛЯЦИОННОМУ ФОРМАТУ

Перечислим ряд организационно-технологических оснований для перевода библиотек ENDF в реляционные СУБД. Приводимые соображения справедливы в отношении всех библиотек ядерно-физических данных.

1. Снимаются ограничения по применяемым физическим моделям, типам и объемам данных:

– можно разместить в одной таблице различные версии данных, что упрощает их сравнение и использование в сопоставительных расчетах;

– структура данных не накладывает ограничений на алгоритмы обработки;

– хранить данные можно одновременно как текст (для обеспечения и проверки идентичности с первичными источниками и аналогичными библиотеками) и как числовую информацию – для обработки и передачи в приложения.

2. Минимизируются трудозатраты на манипуляции с данными в силу исключения программирования для решения некоторых типичных задач:

– табличная форма обеспечивает возможность визуального контроля корректности данных;

– различные корреляции и зависимости могут выявляться оценщиком путем оперативного вывода данных на график с возможностью выбора интервалов и различных столбцов таблицы в качестве аргументов или функций;

– реляционные таблицы могут непосредственно импортироваться в другие приложения, например, в статистические пакеты.

3. Минимизируется время ввода данных и время доступа к данным:

– пользователь практически мгновенно может перейти к нужным строкам таблицы или получить интересующую его выборку или группировку данных;

– бумажная таблица может быть отсканирована и импортирована в базу данных с сохранением исходной структуры за несколько минут;

– добавление записей (например, обусловленное детализацией энергетического интервала) или столбцов (например, столбца погрешностей констант) является стандартной операцией, не требующей реструктурирования файла путем изменения или добавления какой-либо управляющей информации.

4. Упрощается процесс обучения и снижаются требования к квалификации пользователей для работы с массивом данных:

– реляционная БД выполняет обучающую функцию, предоставляя доступ к таблично организованным данным в содержательных терминах;

– работа с данными идет на естественном языке предметной области без использования характерного для формата ENDF специфического синтаксиса на основе указателей наличия и типа данных;

– большая часть задач по анализу данных, в том числе вычислительного характера, решается посредством «запросов по образцу», т.е. декларативным указанием на результат, который следует найти или вычислить, а не описанием алгоритма поиска или вычисления;

– хранение данных в РБД исключает необходимость явного размещения и контроля управляющей информации (разделительных записей, параметров навигации,

указателей типа данных, счетчиков, флагов и т.д.). Тем самым снимается задача разработки соответствующих руководств по организации и форматированию данных, не тратится время на изучение управляющей информации, на освоение ее корректного применения в обрабатываемых программах.

5. Упрощается процесс разработки приложений:

- ряд задач не требует программирования и решается в диалоговом режиме;
- исключается программирование процедур выделения памяти, перемещения данных между внешней и оперативной памятью, а также процедур графического вывода;
- выборка данных выполняется по содержательным именам;
- последовательность операций, выполняемых в диалоговом режиме, может быть записана и сохранена как макрос для повторного исполнения или включена в состав создаваемого приложения;
- добавление в таблицу новых столбцов не требует, как правило, изменения модулей запросов, осуществлявших выборку данных из таблицы первоначальной структуры.

6. Упрощается процесс верификации данных и валидации приложений:

- корректность данных проверяется автоматически в момент их ввода в таблицу;
- упрощается визуальный контроль данных;
- вновь разрабатываемые приложения не требуют разработки и отладки специфических процедур доступа к данным.

7. Упрощается взаимодействие с другими приложениями на основе стандартных решений, ориентированных на применение реляционных СУБД:

- удаленный доступ к данным и вычислительным ресурсам обеспечивается стандартными технологиями клиент-сервер (ADO, ADO.NET, RMI и др.);
- исключается разработка специфических интерфейсов при создании программ реакторных расчетов, интегрированных с пакетами CAD/CAE;
- становятся доступны для анализа данных технологии DATA MINING.

Разработать на языке реляционной СУБД модуль генерации и верификации файлов исходных данных для той или иной расчетной программы можно более быстро и эффективно, чем средствами традиционных языков программирования. Соответственно, результаты расчета, сохраняемые в таблицах некоторой СУБД, непосредственно пригодны для визуализации, статистической обработки и передачи в другие приложения.

ОТОБРАЖЕНИЕ ДАННЫХ ENDF В РЕЛЯЦИОННЫЙ ФОРМАТ

Для рассматриваемой предметной области – ядерные реакции и распады – формат ENDF [14, 15] уже предоставляет формализованную модель в терминах логических указателей и физических параметров, для размещения которых отводится 10 полей текстовой строки:

P1, P2, P3, P4, P5, P6, Материал, Файл, Секция, Строка. (1)

Задача, таким образом, состоит в следующем.

1. Представить данные ENDF посредством инфологических формул вида

Сущность [Атрибут_1, Атрибут_2, ..., Атрибут_N], (2)

описывающих предметную область в терминах объектов, процессов, явлений (сущности) и их свойств (атрибуты). Инфологическая формула задает структуру (содержание строки) реляционной таблицы, в которой столбцы представляют атрибуты некоторой сущности, а строки – ее экземпляры, различающиеся значениями атрибутов. Совокупность инфологических формул (2) полностью определяет схему базы данных – перечень взаимосвязанных таблиц. Связи таблиц задаются посредством ключей – одно-

именных атрибутов, присутствующих в разных таблицах.

2. Отобразить строки (1) в записи реляционных таблиц (2). Такая конвертация не является элементарной операцией, поскольку в файлах ENDF нет указателей начала и конца блоков данных определенного смыслового содержания. Кроме того, алгоритм конвертации будет определяться, очевидно, принятой схемой базы данных.

Стандартную строку (1) файла ENDF можно интерпретировать как запись реляционной таблицы из 10-ти столбцов. При этом нарушается принцип однородности данных в столбце реляционной таблицы, но использование таких промежуточных рабочих таблиц, в которые предварительно импортируются файлы ENDF, упрощает процедуру конвертации. В частности, заголовочные строки секций и подсекций конкретного файла ENDF содержат однородные списки атрибутов. Поэтому таблицы, собранные из таких строк, уже находятся в первой нормальной форме [16 – 19]. Сохранение метаданных в этих таблицах обеспечивает верификацию и валидацию как исходных файлов ENDF, так и процедур конвертации.

В окончательной версии реляционной базы метаданные ENDF исключаются. Структура таблицы, т.е. перечень ее полей (столбцов), определяется только природой конкретной физической информации. В столбце реляционной таблицы хранятся однородные данные, например, полное сечение, и каждая запись (строка таблицы) идентифицируется (в данном случае) соответствующим значением энергии. Требование однородности данных в отдельных столбцах обуславливает выделение в рассматриваемой предметной области следующих основных сущностей: материалы, взаимодействия (реакции), распады и распределения. Под распределениями здесь понимаются любые функциональные зависимости. Вся остальная информация может трактоваться как совокупность свойств (атрибутов) указанных сущностей.

Все единичные (скалярные) свойства однотипных объектов могут быть сохранены в одной таблице. Множественные свойства выражаются, как правило, распределениями – функциональными зависимостями (от температуры, энергии, массы, орбитального момента и т.д.). В таблице функциональную зависимость можно размещать или «по строке», или «по столбцу». Оба представления имеют свои удобства. Строковое размещение обеспечивает оперативность выборки для размещения данных в массивах. Для столбцов система управления предусматривает обычно возможность вывода на график в диалоговом режиме.

При строковом размещении распределений значения одного из аргументов берутся в качестве имен столбцов. Строка содержит все значения распределения для конкретного материала, реакции и текущей комбинации остальных аргументов. Добавив атрибут [*Тип распределения*], можно все распределения (например, энергетические) сохранить в одной таблице. Существенным ограничением здесь является недостаточность числа столбцов, которое СУБД может сопоставить имеющимся точкам аргумента. Кроме того, для файлов оцененных данных типична ситуация, когда множества точек аргумента для различных распределений не совпадают и используются разные схемы интерполяции. Строковое размещение этой информации возможно, но нарушает принцип однородности данных в столбце.

При столбцовом размещении функция и (скалярный) аргумент рассматриваются как атрибуты, значениям которых отводится два столбца. Необходимо также предусмотреть столбцы [*Интервал интерполяции*], [*Тип интерполяции*]. Все пары столбцов [*Аргумент*], [*Функция*] можно разместить в одной таблице со строками вида «материал - реакция - все сечения» или «реакция - сечение - все материалы». В диалоговом режиме работы с таблицей это ускоряет выбор столбцов для вывода на график или для вычисления интерполируемых значений. Этим оправдывается наличие относительно большого числа пустых полей в таблицах описанной организации.

Очевидна иерархическая структура ядерно-физических данных, обусловленная тем, что определенному значению некоторого параметра соответствует, как правило, кортеж значений детализирующих величин. Совокупность кортежей естественно объединить в таблицу, в которой каждая запись (кортеж) идентифицируется соответствующим значением указанного параметра. Если при данном значении параметра (для данного кортежа значений) реализуется некоторое множественное свойство (например, специфическая энергетическая зависимость), то оно выражает отношение «один ко многим» и порождает таблицу следующего уровня. Чтобы связать эту таблицу с детализируемым параметром, значения последнего естественно перенумеровать, добавив соответствующий столбец, и составить название таблицы из названия параметра и номера его значения. Обычно такие таблицы объединяются в одну, содержащую ключевое поле с номером параметра.

Нумерующие (индексные) столбцы являются простым и удобным средством для навигации и поиска в иерархии таблиц. Однако такая иерархия может быть очень глубокой, и количество индексных столбцов окажется сопоставимо с количеством столбцов, хранящих содержательную информацию. В таком случае уместно применить для свертки индексов технику нумерующих функций, аналогичных нумерующей функции $ZA = 1000 * Z + A$ (Z – зарядовое число, A – массовое число материала), используемой в формате ENDF для именования материалов. Например, эту функцию можно интерпретировать как запись числа в системе счисления с основанием $P = 10^3$, которое заведомо превосходит объемы возможных значений в индексных столбцах и позволяет объединить кортеж значений M ключевых полей в единственное адресное число

$$N = \sum_{m=1}^M P^{m-1} e_m.$$

Восстановление исходного кортежа индексов выполняется по известным алгоритмам [20].

ИНФОЛОГИЧЕСКАЯ МОДЕЛЬ ФАЙЛА СЕЧЕНИЙ

Файл ENDF MF=3 содержит только сечения реакций и производные величины в виде функции от энергии падающей частицы. Эти данные определяют простую инфологическую модель, содержащую три сущности: Материал, Ядерная реакция, Сечение. Материал MAT описывается двумя единичными атрибутами – зарядовое и массовое числа ZA, AWR в первой заголовочной строке секции файла MF = 3:

```
[MAT, 3, MT/ ZA, AWR, 0, 0, 0, 0] HEAD
[MAT, 3, MT/ QM, QI, 0, LR, NR, NP/ Eint / σ(E)] TABL
[MAT, 3, 0/ 0.0, 0.0, 0, 0, 0, 0] SEND
```

Единичными атрибутами реакции MT здесь являются материал MAT, разности масс исходных и конечных продуктов QM, QI, характеристика ядра-продукта LR. Множественные атрибуты – это начальная энергия (энергия падающей частицы) E , значение сечения $\sigma(E)$, интервал интерполяции, правило интерполяции. Скалярные атрибуты помещаются в одну таблицу, которая связана по ключевым (идентифицирующим) столбцам [Реакция], [Материал] с таблицей множественных свойств, содержащей столбцы [Энергия падающей частицы], [Сечение]. В этой таблице в столбцах [Интервал интерполяции], [Тип интерполяции] для каждой энергетической точки указывается номер интервала, содержащего данную точку, и тип интерполяции в этом интервале. Таким образом, инфологическая модель задает схему базы данных для файла MF=3, содержащую три таблицы, структура которых описывается следующей

ми формулами:

Материалы [MAT, ZA, AWR],

Реакции [MT, MAT, QM, QI, LR, NR, NP],

Сечения [MT, MAT, E, $\sigma(E)$, Интервал интерполяции, Тип интерполяции].

(Здесь и далее в инфологических формулах для обозначения атрибутов, предусмотренных в библиотеках ENDF, используются стандартные идентификаторы формата ENDF.)

Атрибуты NR, NP, фигурирующие во второй заголовочной строке секции файла MF=3, относятся к метаданным и не включаются в окончательную версию базы данных ядерных сечений. Таблицы находятся в третьей нормальной форме и, согласно известным принципам нормализации реляционных баз данных [16 – 19], в дальнейшей реструктуризации не нуждаются.

ИНФОЛОГИЧЕСКАЯ МОДЕЛЬ ФАЙЛА РАСПРЕДЕЛЕНИЙ ВТОРИЧНЫХ НЕЙТРОНОВ

Файл ENDF MF=5 содержит функции плотностей вероятностей распределения вторичных нейтронов по энергии в реакциях с нейтронами и при делении. Основным информационным блоком здесь является подсекция, хранящая параметры конкретного парциального распределения для ряда значений начальной энергии E . В силу различия наборов атрибутов в классах распределений уместно рассматривать в качестве сущностей именно эти классы. В таком случае экземпляр сущности – это распределение определенного класса при данном значении начальной энергии E .

Все распределения характеризуются вероятностью реализации при конкретном значении начальной энергии E для данной реакции. Кроме того, некоторые функции плотностей вероятностей содержат параметры, также зависящие от начальной энергии падающей частицы. Все указанные данные удобно сохранять в единственной таблице, если они заданы на одной энергетической сетке. В противном случае для каждого параметра необходима отдельная таблица. В частности, для вероятностей реализации распределений вводится таблица следующей структуры:

MF5_P(E) [*Реакция, Материал, Тип распределения, Начальная энергия, Вероятность реализации, Интервал интерполяции, Тип интерполяции*].

Остальные параметры, зависящие от начальной энергии, – температура ядра, параметры спектра Уатта и т.д., размещаются в таблицах аналогичной структуры.

Единичным атрибутом ряда распределений является константа U , определяющая верхний предел энергии вторичных нейтронов, $0 \leq E^* \leq (E - U)$. Эту константу уместно сохранить в таблице

MF5_U [*Реакция, Материал, Тип распределения, U*].

Для распределения типа LF = 1 задается набор значений начальной энергии E и таблица значения вероятностей $g(E \rightarrow E')$ как функции энергии вторичных нейтронов E^* , причем предусматривается возможность интерполяции как по энергии E , так и по энергии E^* . Очевидно, эти данные можно хранить в одной таблице следующей структуры:

MF5_LF1 [*Реакция, Материал, Начальная энергия, Энергия вторичных нейтронов E^* , Интервал интерполяции E^* , Тип интерполяции E^* , Вероятность*].

Для исключения дублирования параметры интерполяции по начальной энергии

выделяются в отдельную таблицу:

MF5_LF1_Интерполяция_по_начальной_энергии [*Реакция, Материал, Начальная энергия, Интервал интерполяции, Тип интерполяции*].

Аналогичные таблицы параметров интерполяции предусмотрены для всех типов распределений.

Спектр испарения общего вида (LF=5) задается совокупностью значений температуры ядра $Q(E)$ в зависимости от начальной энергии E и совокупностью значений функции плотности вероятности $g(x)$, где $x = E^*/\Theta(E)$. Это обуславливает использование двух таблиц, структура которых описывается следующими формулами:

MF5_LF5_TETA [*Реакция, Материал, Начальная энергия, $\Theta(E)$, Интервал интерполяции E , Тип интерполяции E*],

MF5_LF5_G [*Реакция, Материал, x , $g(x)$, Интервал интерполяции, Тип интерполяции*].

Для спектра Максвелла (простой спектр деления, LF=7) и спектра испарения (LF=9) в файле ENDF хранится только распределение температуры ядра $\Theta(E)$, так что для каждого спектра можно обойтись одной таблицей со структурой, подобной таблице MF5_LF5_TETA. Три таблицы аналогичной структуры используются также для хранения параметров $a(E)$, $b(E)$ спектра Уатта (LF=11) и максимальной температуры ядра $T_M(E)$, фигурирующей в формуле спектра Мэдланда и Никса (энергозависимый спектр нейтронов деления, LF=12).

ИНФОЛОГИЧЕСКАЯ МОДЕЛЬ ЭНЕРГОУГЛОВЫХ РАСПРЕДЕЛЕНИЙ ПРОДУКТОВ

Файл ENDF MF=6 хранит данные об энергоугловых распределениях нейтронов, фотонов, заряженных частиц и остаточных ядер, появляющихся как продукты ядерных реакций. Продукт характеризуется сечением образования:

$$\sigma_i(\mu, E, E^*) = \sigma(E) * y_i(E) f_i(\mu, E, E^*) / 2\pi . \quad (3)$$

Эта формула определяет возможную инфологическую модель на основе трех сущностей:

Сечение_реакции [*Материал, Реакция, Начальная энергия E , $\sigma(E)$*],

Выход_продукта [*Материал, Реакция, Продукт, Начальная энергия E , $y(E)$*],

Энергоугловое_распределение [*Материал, Реакция, Продукт, Начальная энергия E , Энергия продукта E^* , μ , $f(E, E^*, \mu)$*].

В таком виде может быть представлена окончательная версия базы данных по энергоугловым распределениям, хранящая значения распределений на достаточно мелкой сетке. Последнее позволяет заменить вычисления выборкой или по умолчанию использовать линейную интерполяцию на всех интервалах изменения аргументов.

На практике число сущностей и соответствующих таблиц увеличивается для хранения параметров распределений в различных представлениях. Единичные атрибуты продукта выбираются из заголовочной строки секции файла MF=6:

Продукт [*Материал, Реакция, Продукт, ZAP, LIP, AWP, LAW, LCT*]. (4)

В силу многообразия дополнительных признаков продукта его стандартные идентификаторы [ZAP, LIP], принятые в формате ENDF, сохраняются в таком качестве в окончательной версии базы данных наряду с развернутым названием. Данные таблицы (4) достаточны для описания энергоугловых распределений при изотропном

испускании частиц с дискретными значениями энергии ($LAW=3$), а также при упругом рассеянии заряженных частиц ($LAW=5$).

Непрерывное энергоугловое распределение ($LAW=1$) характеризуется двумя единичными атрибутами [LANG, LEP], для которых отводится таблица

MF6_LAW1_Параметры [*Материал, Реакция, Продукт, ZAP, LIP, Способ задания углового распределения LANG, Указатель интерполяционной схемы для энергии продукта реакции LEP*].

Параметры интерполяции по начальной энергии выделяются в отдельную таблицу:

MF6_LAW1_Интерполяция_по_начальной_энергии [*Реакция, Материал, Начальная энергия, Интервал интерполяции, Тип интерполяции*].

Энергоугловые зависимости в данном случае могут храниться в таблице

MF6_LAW1_Распределения_в_столбце [*Материал, Реакция, Продукт, ZAP, LIP, Начальная энергия E, Энергия продукта E*, Параметр распределения b(E, E*)*], (5)

где параметры распределения интерпретируются согласно значению указателя LANG. Однако в силу ограниченного числа параметров в используемых представлениях (разложение по полиномам Лежандра или систематика Калбаха-Манна) уместно каждый параметр, зависящий от энергий E, E^* , интерпретировать как отдельный атрибут.

В таком случае структура таблицы принимает вид

MF6_LAW1_Распределения_в_строке [*Материал, Реакция, Продукт, ZAP, LIP, Начальная энергия E, Энергия продукта E*, NA, b₀(E, E*), b₁(E, E*), ..., b_{NA}(E, E*)*]. (6)

Этот вид уменьшает дублирование значений в полях [*Начальная энергия E, Энергия продукта E**]. Сохранение атрибута NA – количество угловых параметров – упрощает обработку массива, в который копируется строка параметров.

Данные для описания энергоугловых распределений в соответствии с законами кинематики рассеяния двух тел ($LAW=2$) можно представить инфологическими формулами, близкими к формулам (5), (6). Здесь только отсутствует зависимость от энергии продукта:

MF6_LAW2_Распределения_в_столбце [*Материал, Реакция, Продукт, ZAP, LIP, Начальная энергия E, Параметр распределения b(E)*],

MF6_LAW2_Распределения_в_строке [*Материал, Реакция, Продукт, ZAP, LIP, Начальная энергия E, NA, b₀(E), b₁(E), ..., b_{NA}(E)*].

Аналогичным образом строятся инфологические формулы для других файлов из библиотек ENDF.

ЗАКЛЮЧЕНИЕ

Отмеченная в [15] большая трудоемкость перевода библиотек ENDF в реляционные базы данных может быть существенно уменьшена в рамках следующей технологии:

- построение адекватных инфологических формул для файлов ENDF;
- использование процедур конвертации, написанных на встроенных языках СУБД;
- выполнение конвертации непосредственно в среде СУБД, куда предварительно построчно копируются файлы ENDF.

Построение инфологических моделей файлов ENDF выполняется достаточно оче-

видным образом, что подтверждают приведенные в работе примеры. Инфологические формулы целесообразно записывать непосредственно по определяющим соотношениям вида (3). В результате задаются, как правило, простая структура реляционных таблиц и семантически ясная схема базы данных. При необходимости инфологические формулы содержат атрибуты метаданных, присутствующие в исходных информационных массивах. В частности, это могут быть все атрибуты, собранные в заголовочных строках секций и подсекций файлов ENDF. Эти атрибуты определяют перечень столбцов реляционных таблиц, хранящих значения скалярных свойств материалов и реакций, а также соответствующий перечень таблиц для скалярных параметров специфических классов энергоугловых распределений. Основной сущностью, требующей специальной таблицы, является то или иное параметрическое представление некоторого множественного атрибута – функции, описывающей зависимость от скалярного или векторного аргумента. Соответственно, экземпляр сущности – это кортеж значений параметров функции при данном значении аргумента. Реакция, материал, продукт реакции и аргумент функции при этом выступают как ключевые атрибуты, идентифицирующие экземпляр сущности и обеспечивающие связь данных в различных таблицах.

Текущий массив инфологических моделей для библиотек ENDF размещен на общедоступных сайтах <http://178.215.91.20/dn>, <ftp://178.215.91.20/dn>.

Литература

1. Юферов А.Г. Конвертация библиотеки РОСФОНД-А в реляционную базу данных. Препринт ФЭИ-3194. – Обнинск: ГНЦ РФ-ФЭИ, 2011. – 28 с.
2. Юферов А.Г., Комиссаров С.А., Николаев М.А. Инфологическая модель файла резонансных параметров. Препринт ФЭИ-3233. – Обнинск: ГНЦ РФ-ФЭИ, 2013. – 40 с.
3. Юферов А.Г. Николаев М.А., Комиссаров С.А. Верификация и валидация файла резонансных параметров в реляционном формате. Препринт ФЭИ-3236. – Обнинск: ГНЦ РФ-ФЭИ, 2013. – 20 с.
4. Варламов В.В., Ишханов Б.С. Базы ядерных данных в научных исследованиях. Электронный ресурс: <http://nuclphys.sinp.msu.ru/bd/index.html> (дата доступа 13.10.2017).
5. Варламов В.В., Гончарова Н.Г., Ишханов Б.С. Физика ядра и банки ядерных данных. Электронный ресурс: <http://nuclphys.sinp.msu.ru/ndb/index.html> (дата доступа 13.10.2017).
6. Митропольский И.А. Структура атомного ядра. Веб-курс лекций. Электронный ресурс: http://nsp.phys.spbu.ru/images/Kursi_lectsiy/Mitropolskiy/Structura_atomnogo_yadra/LO.pdf (дата доступа 13.10.2017).
7. Метаданные для информации научного профиля. Электронный ресурс: http://www.elbib.ru/index.phtml?env_page=methodology/metadata/md_review/md_descrip_scientific.html (дата доступа 13.10.2017).
8. Gray J., Liu D. T., Nieto-Santisteban M. A., Szalay A., DeWitt D. J., Heber G. Scientific data management in the coming decade, SIGMOD Record. – Dec 2005. – Vol. 34. – No. 4. – PP. 34-41.
9. Бартунов О.С., Велихов П.Е. Научные вызовы технологиям СУБД. // Суперкомпьютеры. – 2011. – № 5. – С. 28-31.
10. Пляскин В.И., Косилов Р.А. Справочно-информационные интерактивные системы ядерно-физических данных для различных приложений ядерной физики. – М.: Энергоатомиздат, 2002. – 375 с.
11. Sinitza V. V., Rineiskiy A. A. GRUCON – A Package of Applied Computer Programs, Rep. INDC(СССР)-344. IAEA, Vienna, 1993.
12. Синица В.В. REBUS: интерактивный редактор данных по сечениям в формате ENDF. Препринт ФЭИ-2904. – Обнинск: ГНЦ РФ-ФЭИ, 2002. – 42 с.
13. Блохин П.А., Митенкова Е.Ф. Описание программного комплекса GrayS для формирования γ -источника. Препринт ИБРАЭ-2012-06. – М.: ИБРАЭ РАН, 2012. – 28 с.

14. *Drake M.K.*, Ed. Data Formats and Procedures for the ENDF Neutron Cross Section Library. BNL-50274, 1970. Электронный ресурс: https://digital.library.unt.edu/ark:/67531/metadc67248/m2/1/high_res_d/metadc67248.pdf (дата доступа 13.10.2017).
15. ENDF-6 Formats Manual, Data Formats and Procedures for the Evaluated Nuclear Data File ENDF/B-VI and ENDF/B-VII, BNL-90365-2009, edited by M.W. Herman and A. Trkov (June 2009), revised by M.W. Herman, A. Trkov and D.A. Brown (Dec. 2011).
16. *Date C.J.* An Introduction to Database Systems (8 ed.). Pearson/Addison Wesley, 2004. – 1024 p.
17. *Gray P.M.D.* Logic, algebra, and databases. Chichester, West Sussex, England: E. Horwood; New York : Distributor, Halsted Press, 1984. – 294 p.
18. *Martin J.* Computer Database Organization. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1977. – 712 p.
19. *Maier D.* Theory of Relational Databases. Computer Science Press, 1983. – 652 p.
20. *Knuth D.* The Art of Computer Programming. 2: Seminumerical Algorithms (3rd ed.), Addison-Wesley Professional, 1997. – 784 p.

Поступила в редакцию 29.11.2017

Автор

Юферов Анатолий Геннадьевич, доцент, канд. физ.-мат. наук
E-mail: anatoliy.yuferov@mail.ru

UDC 621.039.51:539.17:004.652.4

INFOLOGICAL MODELS OF THE ENDF-FORMAT NUCLEAR DATA

A.G. Yuferov

Obninsk Institute for Nuclear Power Engineering, NRNU MEPhI
1 Studgorodok, Obninsk, Kaluga Reg., 1249040, Russia

ABSTRACT

Issues of the infologic modeling of the ENDF-format nuclear data libraries for the purpose of converting ENDF files into a relational database are considered. The transfer to a relational format will make it possible to use standard public software for nuclear data processing, which simplifies the conversion and operation of this information array. Infological models are described using formulas of the «Entity (List of Attributes)» type. The proposed infological formulas are based on the physical nature of data and defining theoretical relations. This eliminates the need for a special notation to be introduced to describe the structure and the content of data, which, in turn, facilitates the use of relational formats in design codes and solution of nuclear data estimation problems. The concept of nuclear informatics has been formulated based on relational DBMS technologies as one of the tools to solve the problem of «big data» in modern science and technology. The organizational and technological grounds for the transfer of ENDF libraries to a relational format are presented. The requirements to the nuclear data presentation formats supported by relational DBMS are listed. Features of the infological model construction, conditioned by the hierarchical nature of nuclear data, are identified. The order for the ENDF metadata preservation is presented, which can be useful for the verification and validation (testing of the structural and syntactical validity and operability) of both source data and the procedures for conversion into a relational format. Formulas of infological models are presented for the cross-sections file, the secondary neutron energy distributions file, and the nuclear reaction product energy-angular distributions file. A complete array of infological models for ENDF libraries and the generation modules of respective relational tables have been published on a publicly available website.

Key words: ENDF libraries, infologic modeling, relational format.

REFERENCES

1. Yuferov A.G. *Converting the ROSFOND library to a relational database*. IPPE Preprint-3194. Obninsk. FEI Publ., 2011. 28 p. (in Russian).
2. Yuferov A.G., Komissarov S.A., Nikolaev M.A. Relational model of the file of angular distributions: IPPE Preprint-3235. Obninsk. FEI Publ., 2013. 20 p. (in Russian).
3. Yuferov A.G. Nikolaev M.A., Komissarov S.A. *Verification and validation of the file of resonant parameters in the relational format*. IPPE Preprint-2828. Obninsk. FEI Publ., 2013. 20 p. (in Russian).
4. Varlamov V.V., Ishkhanov B.S. Nuclear databases in scientific research. Available at: <http://nuclphys.sinp.msu.ru/bd/index.html> (accessed Oct. 13, 2017) (in Russian).
5. Varlamov V.V., Goncharova N.G., Ishkhanov B.S. Physics of nuclei and nuclear data banks. Available at: <http://nuclphys.sinp.msu.ru/ndb/index.html> (accessed Oct. 13, 2017) (in Russian).
6. Mitropolskiy I.A. The structure of the atomic nucleus. Web lectures. Available at: http://nsp.phys.spbu.ru/images/Kursi_lectsiy/Mitropolskiy/Structura_atom_nogo_yadra/L0.pdf (accessed 13 Oct. 2017) (in Russian).
7. Metadata for information scientific profile. Available at: http://www.elbib.ru/index.phtml?env_page=methodology/metadata/md_review/md_descrip_scientific.html (accessed Oct. 13, 2017) (in Russian).
8. Gray J., Liu D.T., Nieto-Santisteban M.A., Szalay A., DeWitt D.J., Heber G. Scientific data management in the coming decade, SIGMOD Record. Dec 2005, v. 34, no. 4, pp. 34-41.
9. Bartunov O.S., Velikhov P.E. Scientific challenges to DBMS technology. *Superkompyutery*, 2011, no. 5, pp. 28-31 (in Russian).
10. Plyaskin V.I., Kosilov R.A. *Reference-information interactive systems of nuclear physical data for various applications of nuclear physics*. Moscow. Energoatomizdat Publ., 2002. 375 p. (in Russian).
11. Sinita V.V., Rineiskiy A.A. GRUCON – A Package of Applied Computer Programs. Rep. INDC(CCCP)-344. IAEA, Vienna, 1993.
12. Sinita V.V. *REBUS: an interactive data editor for nuclear cross sections in the ENDF format*. Obninsk, FEI Publ., 2002. 42 p. (in Russian).
13. Blokhin P.A., Mitenkova E.F. *Specification of GrayS program for g-source generation*. Preprint IBRAE-2012-06. Moscow. NSI RAS Publ., 2012. 28 p. (in Russian).
14. Drake M.K., Ed. Data Formats and Procedures for the ENDF Neutron Cross Section Library. BNL-50274, 1970. Available at: https://digital.library.unt.edu/ark:/67531/metadc67248/m2/1/high_res_d/metadc67248.pdf (accessed Oct. 13, 2017).
15. ENDF-6 Formats Manual, Data Formats and Procedures for the Evaluated Nuclear Data File ENDF/B-VI and ENDF/B-VII, BNL-90365-2009, edited by M.W. Herman and A. Trkov (June 2009), revised by M.W. Herman, A. Trkov and D.A. Brown (Dec. 2011).
16. Date C.J. *An Introduction to Database Systems (8-th ed.)*. Pearson/Addison Wesley, 2004. 1024 p.
17. Gray P.M.D. *Logic, algebra, and databases*. Chichester, West Sussex, England: E. Horwood; New York: Distributor, Halsted Press, 1984. 294 p.
18. Martin J. *Computer Database Organization*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1977. 712 p.
19. Maier D. *Theory of Relational Databases*. Computer Science Press, 1983. 652 p.
20. Knuth D. *The Art of Computer Programming. 2: Seminumerical Algorithms (3-rd ed.)*, Addison-Wesley Professional, 1997. 784 p.

Author

YuferovAnatoliy Gennadyevich, Assistant Professor, Cand. Sci. (Phys.-Math.)
E-mail: anatoliy.yuferov@mail.ru