

УДК 681.3.01

ПРЕОБРАЗОВАНИЕ ДОКУМЕНТАЦИИ БИБЛИОТЕКИ ЧИСЛЕННОГО АНАЛИЗА В ФОРМАТ HTML

Н. И. Волченскова¹, И. В. Зюряева¹, Т. В. Ламзина¹

Рассматриваются средства для автоматизированного представления пользовательской документации по Библиотеке программ решения типовых задач численного анализа в HTML-документы. Обсуждаются способы преобразования структурных разделов исходных документов и приводятся соответствующие шаблоны. Излагаются рекомендации по работе с программой преобразования.

Ключевые слова: библиотеки стандартных программ, численный анализ, численные методы, HTML-документы, гипертекстовые документы, программная документация.

1. Введение. В связи с проведением работ по представлению Руководства по использованию Библиотеки численного анализа (БЧА) НИВЦ МГУ [1, 2] в сети Интернет возникла задача автоматизации процессов преобразования файлов документации из текстового представления в гипертекстовые документы (HTML-документы) [3, 4] с соответствующей перекодировкой исходных ASCII-файлов в коды WINDOWS-CP1251. При этом надо было учитывать то обстоятельство, что исходные тексты создавались различными людьми в различных редакторах (Norton, Chiwriter и др.), имели свою предысторию создания и могли отличаться кодировкой базовых элементов (например, использование русских (латинских) букв, совпадающих по начертанию с латинскими (русскими) в одном слове заголовка, подчеркивание (надчеркивание) заголовков и т.п.).

Основной объем документации БЧА составляют файлы описаний подпрограмм, составленные в соответствии с общими правилами. Автоматизация преобразования каждого такого файла в HTML-документ основывается на самой структуре входного документа, состоящего из десяти фиксированных структурированных разделов. Вид раздела определяется по его заголовку, состоящему из двух строк — имени заголовка раздела и строки подчеркивания (или надчеркивания). Каждый раздел может содержать только информацию определенного вида и структуры, что определяет алгоритм преобразования. Преобразование близких по структуре разделов осуществляется одним блоком разработанной авторами программы-конвертора и основано на использовании шаблонов языка HTML, которые в выходном файле перемежаются со строками исходного текста.

В последующих разделах данной работы будут рассмотрены шаблоны для создания HTML-документа, исследованы особенности преобразования отдельных разделов исходного текста и описаны правила работы с программой-конвертором.

2. Шаблоны для создания HTML-документа. Шаблоны языка HTML представляют из себя отдельные теги или группу тегов, определенных в этом языке. Каждый шаблон может содержать либо только элементы языка HTML, либо в него в качестве параметров может подаваться информация из исходного текста, которая встраивается в шаблон. Обращение к шаблону обеспечивает запись строк шаблона в выходной файл. Шаблоны подразделяются на несколько видов: шаблон начала файла, шаблон конца файла, шаблоны для заголовков разделов, шаблоны для формирования выходного структурированного текста (абзацев, таблиц, подтаблиц, преформатированного текста).

Процесс преобразования включает в себя следующие шаблоны языка HTML:

BeginBody	- начало файла HTML
EndBody	- конец файла HTML
ProgName	- заголовок "Подпрограмма"
FuncName	- заголовок "Функция"
Naznach	- заголовок "Назначение"
MathDiscrip	- заголовок "Математическое описание"
Use	- заголовок "Использование"
Operator	- оператор в разделе "Использование"
Params1	- заголовок "Параметры: нет"

¹ Научно-исследовательский вычислительный центр, Московский государственный университет им. М.В. Ломоносова, 119899, Москва; e-mail: iraz@srcc.msu.su

Params2	- заголовок "Параметры"
Verchin1	- заголовок "Версии: нет"
Verchin2	- заголовок "Версии"
CallProgramm1	- заголовок "Вызываемые подпрограммы: нет"
CallProgramm2	- заголовок "Вызываемые подпрограммы"
Notes1	- заголовок "Замечания по использованию: нет"
Notes2	- заголовок "Замечания по использованию"
Example	- заголовок "Пример использования"
BeginAbz	- начало абзаца
EndAbz	- конец абзаца
BeginPre	- начало преформатированного текста
EndPre	- конец преформатированного текста
BeginTable	- начало таблицы параметров
Table1	- 1-й столбец строки таблицы параметров
Table2	- 2-й столбец строки таблицы параметров
EndTable	- конец таблицы
Sub1Table1	- 1-й столбец строки первой подтаблицы
Sub1Table2	- 2-й столбец строки первой подтаблицы
Sub2Table1	- 1-й столбец строки второй подтаблицы
Sub2Table2	- 2-й столбец строки второй подтаблицы
BeginTableNotes2	- начало таблицы замечаний
TableNotes1	- 1-й столбец строки таблицы замечаний
TableNotes1_1	- 1-й столбец строки таблицы литературы
TableNotes2	- 2-й столбец таблицы замечаний без номеров
TableNotes2_1	- 2-й столбец таблицы замечаний с номерами
TableNotes3	- 3-й столбец таблицы замечаний с номерами

Ниже приводятся примеры шаблонов HTML.

```
#:BeginTable
*
* Начало таблицы параметров
*
*
fwrite_line &fn. \
:<table cellpadding="6" cellspacing="0" width="99%">
return

#:Operator
params_get fn oper
*
* Оператор в разделе 'Использование'
*
fwrite_line &fn. \
:<pre>    &oper. </pre>
return

#:EndAbz
*
* Конечный тег абзаца
*
fwrite_line &fn. \
: </p>
return
```

3. Преобразование отдельных разделов исходного текста. В начале и в конце работы программы-конвертора в выходной файл записываются соответствующие шаблоны начала и конца файла HTML. Выявление заголовка раздела определяет блок программы, предназначенный для дальнейшего анализа исходного текста раздела. Отдельные блоки программы анализируют однотипные структуры, особенности которых рассматриваются ниже.

Первый раздел исходного файла “Подпрограмма” или “Функция” содержит имя файла, которое перекодируется в стандартный вид — прописные латинские буквы (в силу использования различных редакторов и долгой предыстории подготовки исходных текстов имена в исходных файлах могут содержать строчные буквы, русскую кодировку букв и даже цифру 0 вместо буквы O).

Второй и третий разделы исходного файла “Назначение” и “Математическое описание” могут содержать абзацы, формулы, пронумерованный список литературы. Начало абзаца определяется пустой строкой, предшествующей абзацу (начальный отступ абзаца в исходном тексте отсутствует), формула — строкой, начинающейся отступом от левого края раздела, список литературы — цифровой нумерацией в начале абзаца. Если в исходном тексте встречаются переносы слов, то они убираются путем “переброски” начала слова на следующую строку входного текста. Исходный текст формул записывается в выходной файл в преформатированном виде и подлежит, если необходимо, индивидуальному редактированию. Таким образом, в выходной файл заносится информация:

Для абзаца:	Для формул:	Для списка:
BeginAbz	<pre>'1 формула'	BeginTableNotes2
'1 строка абзаца'	...	TableNotes1_1
...	'K формула'	'номер'
'K строка абзаца'	EndPre	TableNotes2
EndAbz		'1 строка без номера'
		...
		'K строка'
		...
		EndTable

Четвертый раздел исходного файла “Использование” включает в себя первый оператор описываемой библиотечной программы, который после перекодировки в прописные латинские буквы и разрядки пробелами списка параметров записывается в выходной файл в преформатированном виде: <pre>'оператор'</pre>. Если оператор имеет вторую строку продолжения, начинающуюся с символа “*”, то символ продолжения перед записью убирается.

Пятый, шестой и седьмой разделы исходного файла “Параметры”, “Версии” и “Вызываемые программы” не являются обязательными и могут отсутствовать, но тогда в документации соответствующий заголовок раздела содержит слово “нет”, которое определяет выбор шаблона для заголовка раздела в выходном файле (со словом “нет” или без него).

Внутренняя структура этих разделов однотипна и представлена в структурированном виде следующим образом:

Левая часть	Символ разделитель	Правая часть
идентификатор	-	текстовый абзац [подтаблица]
ид-тор,ид-тор ...	-	текстовый абзац [подтаблица]
ид-тор,ид-тор, ...	-	текстовый абзац [подтаблица]
ид-тор,ид-тор	-	текстовый абзац [подтаблица]

Подтаблица является необязательной частью раздела и может иметь следующую структуру:

Левая часть	Символ разделитель	Правая часть
ид-тор=N	-	текстовый абзац
ид-тор=N,	-	текстовый абзац
ид-тор>N	-	текстовый абзац
ид-тор>N,	-	текстовый абзац
ид-тор.GE.N	-	текстовый абзац
ид-тор.GE.N,	-	текстовый абзац
ид-тор.NE.N	-	текстовый абзац
ид-тор.NE.N,	-	текстовый абзац
	. . .	
число	-	текстовый абзац

Правая часть отличается многообразием возможных выражений, что не всегда приводит к правильной автоматизированной обработке. Такая структура преобразуется с помощью шаблонов таблиц и подтаблиц и в выходной файл заносится информация:

Основная таблица	1-я подтаблица	2-я подтаблица
BeginTable	BeginTable	BeginTable
Table1	Sub1Table1	Sub2Table1
'идентификатор' -	'выражение' -	'число' -
Table2	Sub1Table2	Sub2Table2
'1 строка абзаца'	'1 строка абзаца'	'1 строка абзаца'
...
'K строка абзаца'	'K строка абзаца'	'K строка абзаца'
...
EndTable	EndTable	EndTable

Если левая часть таблицы исходного текста представляет собой список идентификаторов, то в выходной файл в первый столбец таблицы записывается весь этот список, скорректированный в рамках языка HTML.

При преобразовании раздела "Версии" описание некоторых версий может не включаться в выходной файл (версии "А" и "Е"). Тогда при отсутствии других версий раздел становится пустым, что влечет изменение заголовка раздела (Версии: нет).

Если в разделе встречается текстовый абзац, не представляющий собой явно таблицу или подтаблицу (левая часть отсутствует), то для сохранения структурности в выходной текст этот абзац записывается как соответствующая таблица, но с пустым () первым столбцом ячейки.

Восьмой раздел исходного файла "Замечания по использованию" может не иметь замечаний (что отражается в заголовке), иметь замечания в виде простых абзацев или в виде пронумерованных абзацев. В силу того, что раздел позиционно сдвинут, все виды замечаний целесообразно преобразовать через шаблоны таблиц. Внутри таблицы каждый абзац исходного текста оформляется через шаблоны начала и конца абзаца. Таким образом, в выходной файл заносится следующая информация:

Таблица замечаний без номеров	Таблица замечаний с номерами
BeginTableNotes2	BeginTableNotes2
TableNotes1	TableNotes1
ableNotes2	TableNotes2_1
BeginAbz	'номер'
'1-я строка абзаца'	TableNotes3
...	BeginAbz
'K-я строка абзаца'	'1-я строка абзаца'
EndAbz	...
...	'K-я строка абзаца'
EndTable	EndAbz
	...
	EndTable

Десятый раздел исходного файла "Пример использования" может состоять из строк с операторами Фортрана, строк с результатами счета и строк-комментариев на русском языке. Фортранные строки и строки результата записываются в выходной файл в преформатированном виде после перекодировки в прописные английские символы и разрядки пробелами списков параметров и некоторых операторов и подлежают, если необходимо, индивидуальному редактированию. Строки-комментарии записываются в выходной файл с использованием шаблонов начала и конца абзаца. Таким образом, в выходной файл заносится следующая информация:

Для строк Фортрана и результата	Для строк комментария
<pre>'1-я строка'	BeginAbz
...	'1-я строка'
'K-я строка'	...


```
</body>  
</html>
```

4. Работа с программой преобразования.

Программа преобразования файла документации Библиотеки численного анализа НИВЦ МГУ в HTML-документ (convef.viz) и список шаблонов (ins_html.viz) написаны на языке ВИЗ — Главном управляющем языке базовой технологии создания систем обработки и отображения сложно организованной информации, разработанном в НИВЦ МГУ [5]. Объем программ порядка 39 Кб. Исходные файлы берутся из подкаталогов, в которые затем заносятся сгенерированные HTML-документы с расширением .htm, имеющие имена исходных файлов.

В силу специфики языка ВИЗ двойные кавычки (") исходного текстового файла преобразуются в одинарные кавычки ('), и при необходимости требуется их последующее редактирование.

Программе подается две строки из исходного текстового файла, что дает возможность более точно проанализировать текущую ситуацию и определить последующие действия. Разделы исходной документации должны быть расположены в строго определенном порядке, нарушение которого ведет к ошибке, фиксируемой в диагностическом файле erfil.txt, и прерыванию процесса преобразования текущего файла.

Запуск программы преобразования осуществляется через стартовый файл, содержащий список полных имен входных файлов (ограничения на расширение имен файлов нет). Пример стартового файла на языке ВИЗ приведен ниже.

* Стартовый файл программы преобразования

```
call convef.viz ..\AC\acp1r.nrp  
call convef.viz ..\AS\ask7c.txt  
call convef.viz ..\AS\asc0c.txt  
call convef.viz ..\AF\afb2r.txt  
call convef.viz ..\QS\qsfpr  
return
```

Работа выполнена при поддержке РФФИ, проект № 98-07-90018.

СПИСОК ЛИТЕРАТУРЫ

1. Воеводин В.В., Арушанян О.Б. Структура и организация Библиотеки численного анализа НИВЦ МГУ // Численный анализ на ФОРТРАНе. Вычислительные методы и инструментальные системы. М.: Изд-во Моск. ун-та, 1979.
2. Арушанян О.Б. Автоматизация конструирования библиотек программ. М.: Изд-во Моск. ун-та, 1988.
3. Браун М., Ханикатт Д. HTML 3.2 в подлиннике. Наиболее полное руководство. СПб.: BHV, 1998.
4. Гончаров А. HTML в примерах. СПб.: Питер, 1997.
5. Арушанян О.Б., Богомоллов Н.А., Волченкова Н.И., Ковалев А.Д., Сидицын М.Н. Общее описание Главного управляющего языка Базовой технологии создания систем обработки и отображения сложно организованной информации // Библиотеки и пакеты прикладных программ. М.: Изд-во Моск. ун-та, 1996.

Поступила в редакцию
14.01.2000
