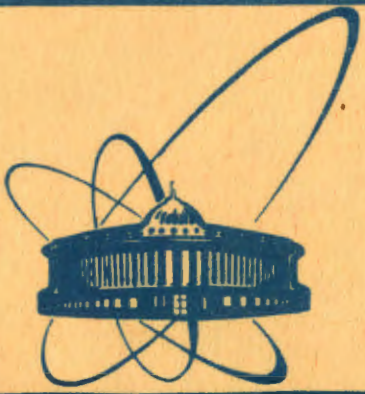

сообщения
объединенного
института
ядерных
исследований
дубна



10-82-463

Т.А.Ершова, С.Г.Олейникова, П.П.Сычев

# ТРАНСПОРТАБЕЛЬНАЯ ПРОГРАММА "ТЕХТА" ДЛЯ ПОДГОТОВКИ ТЕКСТОВОЙ ДОКУМЕНТАЦИИ

1982

При создании программного обеспечения крупных автоматизированных систем как для научных исследований, так и для управления хозяйственной и производственной деятельностью, весьма велик объем связанной с ним документации. Кроме того, документация характеризуется высокой степенью изменяемости и должна точно соответствовать последней версии программного обеспечения. Для подготовки и ведения этой документации логично использовать ЭВМ. Рассматриваемая здесь программа "TEXTA" обеспечивает обработку текстовой информации, подготовку и выдачу ее в книгоподобном виде.

Распространенные в настоящее время аналогичные программы ( BARB /1/ и др.) не вполне удовлетворительны по тем или иным причинам: в частности, из-за уникальности входного языка каждой из этих систем, их нетранспортабельности и, кроме того, первоначальной ориентации на тексты на английском языке.

При создании программы "TEXTA" были поставлены следующие цели:

- обеспечить более высокое "полиграфическое" качество выходного текста, чем это обычно принято, в частности, производить выравнивание текста по краям страницы, автоматический грамматический перенос слова и т.д.;

- достичь возможно более полной транспортабельности программы, т.е. возможности ее переноса на ЭВМ различных типов. Эта задача выполняется путем использования стандартного языка высокого уровня.

В качестве основы при разработке входного языка программы был выбран язык широко распространенной в ОИЯИ аналогичной программы "BARB" . Выбор был сделан по следующим соображениям.

Во-первых, разработка самостоятельного языка сама по себе является довольно сложной задачей и должна быть связана с необходимостью значительного улучшения существующих.

Во-вторых, желательно было обеспечить частичную совместимость из-за большого количества текстов, подготовленных для программы "BARB".

При этом ряд возможностей языка "BARB" исключен, но, в свою очередь, добавлены новые, отсутствующие в "BARB", в частности, автоматический перенос слова, выравнивание текста по правой границе страницы, построение таблиц, режим билистинга.

## 2. Основные возможности

Входные данные для программы готовятся на перфокартах или любым другим образом на файле, записи которого рассматриваются как образы перфокарт. Управление программой осуществляется управляющим символом, находящимся в первой позиции каждой карты. Остальные позиции карты рассматриваются как информационное поле и содержат текстовую информацию. Для некоторых карт информационное поле может содержать различные параметры, управляющие работой программы. Размер информационного поля по умолчанию принимается со 2- й по 72-ю колонку, но может быть изменен при вводе данных.

Рассмотрим основные возможности программы. Текст разбивается на главы, параграфы и разделы. Они автоматически нумеруются, их названия заносятся в оглавление, которое будет напечатано в конце текста. Вывод информации в книгоподобном виде предусматривает разбиение текста на страницы. Пользователь имеет возможность задать необходимые для него параметры страницы, т.е. количество строк в странице, число символов в строке и т.п. Кроме того, возможно размещение двух страниц текста на одном листе АЦПУ (режим билистинга). Смена страницы происходит либо автоматически, либо по требованию пользователя. Каждая страница нумеруется по порядку, в ее заголовок выносятся названия текущей главы и параграфа.

Основным режимом работы программы является накопление текста. Вводимая информация рассматривается как поток "слов". Под "словом" понимается последовательность символов, ограниченная с обеих сторон либо пробелами, либо границей записи. Например, "слово" будет включать знак препинания, если он не отделяется пробелом*. Вводимые одно за другим слова размещаются в текущей строке. В случае, если очередное слово не размещается в строке, рассматривается возможность его грамматического переноса. В программе реализован алгоритм переноса, описанный ниже. Подготовленная строка выравнивается по правому краю страницы путем вставки дополнительных пробелов. В первую очередь дополнительные пробелы вставляются после знаков препинания; если этого недостаточно, то между остальными словами.

* Далее термин "слово" в указанном смысле будет употребляться без кавычек.



Кроме рассмотренных возможностей, программа также позволяет:

- накапливать текст в "поколонном" режиме - при этом страница разбита на ряд столбцов, информация в которые вводится отдельно;
- формировать таблицы;
- делать примечания, которые будут напечатаны внизу соответствующей страницы;
- выделять отдельные фрагменты текста жирным шрифтом;
- печатать абсолютный текст, т.е. переносить данные из карты в строку без каких-либо изменений;
- центрировать текст.

Предусмотрены средства, облегчающие подготовку больших текстовых материалов. Они дают возможность получить достаточно подробную диагностику ошибок (либо на русском, либо на английском языках), вывести на печать параллельно с текстом номера перфокарт, что удобно при устранении ошибок и т.п.

## 3. Особенности реализации программы

В качестве языка программирования был использован стандартный вариант КОБОЛа. КОБОЛ является эффективным и достаточно удобным средством для обработки символьной информации и доступен на большинстве современных ЭВМ. В принципе, при употреблении только стандартных возможностей языка, любая программа на КОБОЛе содержит лишь одну, хорошо выделенную машинно-зависимую часть (так называемый раздел оборудования).

Разумеется, эффективность объектной программы будет при этом зависеть от особенностей данной ЭВМ и используемого компилятора. При написании программы "ТЕХТА" везде, где это было возможно, использовались только стандартные возможности КОБОЛа, независимо от эффективности реализации их на конкретной ЭВМ. В результате программа содержит лишь один (не считая раздела оборудования) машинно-зависимый фрагмент (модуль разбиения слов на слоги), в котором используется полная таблица определенных в данной ЭВМ символов.

Транспортабельность программы подтверждается тем, что она поставлена на ЭВМ CDC-6500 *, ЕС-1060 и ЕС-1040; ее перенос на другие ЭВМ не требует больших усилий.

Программа имеет модульную структуру, ее общая схема представлена на рис.1. Потоки данных на схеме изображены двойными линиями, управление - одинарными. Естественно, что приведены лишь основные

* На ЭВМ CDC-6500 в программу были включены дополнительные возможности в связи с особенностями использования русского шрифта на этой ЭВМ.

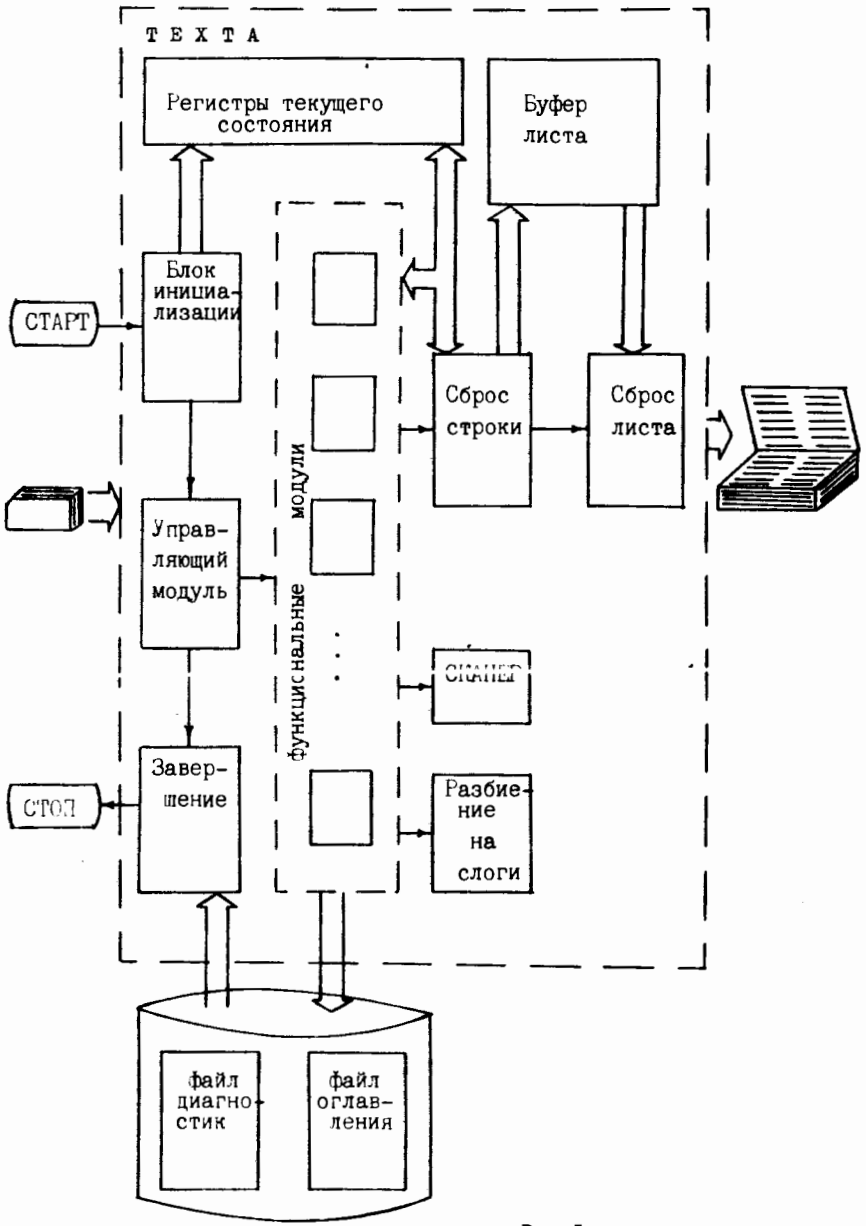


Рис.I.



модули программы и основные потоки информации. Дадим некоторые пояснения к схеме. Очередная запись входных данных считывается управляющим блоком, который распознает управляющий символ и вызывает соответствующий функциональный модуль. Функциональные модули обрабатывают запись входных данных, готовя строки выходного текста. После подготовки очередной строки она сбрасывается в буфер страницы (или листа в режиме билистинга). Заполненный буфер страницы выводится на печать, при этом происходит окончательная доводка текста. В процессе обработки входных данных накапливается файл оглавления и, если затребовано пользователем, файл диагностических сообщений. Оба файла по завершении обработки входных данных считываются модулем завершения, как обычные входные данные, подготовленные по правилам "ТЕХТА" , и выводятся на печать.

Среди вспомогательных модулей выделены два наиболее важных - сканер, перенос слов. Сканер обеспечивает получение очередного слова и используется практически всеми функциональными блоками. Частота его вызовов примерно равна числу слов в тексте. Так как модуль работает значительную часть времени, скорость его выполнения во многом определяет общую производительность программы.

Как правило, программы аналогичного назначения не осуществляют автоматического грамматического переноса слов, хотя некоторые из них производят выравнивание текста по краям страницы. В среднем слова русского языка длиннее английских, поэтому такое решение привело бы к необходимости размещать достаточно большое количество дополнительных пробелов (около 4-6 на строку *), что снижает субъективную оценку качества подготовленного текста на русском языке. Использование рассматриваемого алгоритма переноса уменьшает число лишних пробелов до I-2 на строку*. Модуль переноса вызывается при всяком переполнении очередной строки или столбца в табличном режиме. Частота его вызовов, следовательно, также весьма высока, и в первом приближении равна числу строк в тексте.

Реализован следующий алгоритм переноса слов. Все существующие на данной ЭВМ символы разбиваются на три категории: гласные буквы, согласные буквы и специальные, в число которых включаются цифры и все остальные символы. Рассматриваемое слово заменяется его шаблоном, в котором каждый символ представлен кодом принадлежности его к своей категории. Далее шаблон просматривается слева направо через 4-символьное "окно" на наличие одного из следующих фрагментов:

---

* Разумеется, приведенные цифры весьма приблизительны и зависят от характера текста.

IOOI IOIO OIIO

где О - обозначает гласную, а I - согласную буквы. Нахождение любого из этих фрагментов определяет точку возможного переноса слова:

IOOI ⟶ IO - OI
IOIO ⟶ IO - IO
OIIO ⟶ OI - IO

Разбиение символов на категории производится путем определения таблицы соответствия между внутренним представлением данного символа и его категорией. Эта таблица определяется конкретной ЭВМ и меняется при переносе программы на ЭВМ другого типа. Русские гласные и согласные определены обычным образом, за исключением твердого и мягкого знаков. Из-за их двойственной функции они не могут быть отнесены к гласным или согласным в соответствии с указанными правилами переноса (сравните, например, получаемый перенос слов "ружье" и "мышь" в обоих случаях). В связи с этим они отнесены к специальным символам и, следовательно, слоги с их участием не выделяются. Также к специальным символам отнесены все буквы латинского алфавита, если их внутреннее представление на данной ЭВМ не совпадает с некоторыми буквами русского алфавита.

Несмотря на свою простоту, этот алгоритм работает достаточно хорошо, обеспечивая в подавляющем большинстве случаев правильный перенос слова. Для устранения иногда допускаемых им ошибок требуется гораздо более тщательный анализ структуры слова с выделением не только слогов, но и приставки, корня и т.д. Не говоря уже о сложности поставленной задачи, это привело бы к очень медленной работе модуля.

Общий объем программы составляет около 2,2 тыс. строк исходного текста на КОБОЛе вместе с внутренней документацией. Объем рабочей памяти, необходимый для хранения регистров состояния программы, рабочих переменных, буферов страницы и дополнительных буферов табличного режима и жирной печати составляет около 50 тыс. символов. Размер загрузочного модуля зависит от ЭВМ, версии операционной системы и используемого компилятора. Ниже приведены эксплуатационные характеристики программы для ЭВМ ОИЯИ.

Данные о скорости работы являются ориентировочными и могут значительно меняться в зависимости от характера текста.



Таблица

| ЭВМ | ОС | Компилятор | Размер загрузочного модуля | Средняя скорость работы |
|---|---|---|---|---|
| CDC-6500 | NOS/BE-1 | COBOL | $53276_8$ | 38карт/с |
| CDC-6500 | NOS/BE-1 | COBOL5 | $50226_8$ | 41карт/с |
| ЕС-1060 | ОС ЕС 6.1 | ОС ЕС COBOL | 122 К | 40карт/с |
| ЕС-1060 | OS/360(IBM) | IBM OS COBOL | 122 К | 42карт/с |
| ЕС-1040 | OS/360(IBM) | IBM OS COBOL | 122 К | 21карт/с |
| ЕС-1033 | ОС ЕС 6.1 | ОС ЕС COBOL | 122 К | 10карт/с |

ЛИТЕРАТУРА

1. Gage B. Text Formating Program (BARB). CERN Program Library,Q500 .

Рукопись поступила в издательский отдел
17 июня 1982 года.

## НЕТ ЛИ ПРОБЕЛОВ В ВАШЕЙ БИБЛИОТЕКЕ?

**Вы можете получить по почте перечисленные ниже книги, если они не были заказаны ранее.**

| | | |
|---|---|---|
| Д13-11182 | Труды IX Международного симпозиума по ядерной электронике. Варна, 1977. | 5 р. 00 к. |
| Д17-11490 | Труды Международного симпозиума по избранным проблемам статистической механики. Дубна, 1977. | 6 р. 00 к. |
| Д6-11574 | Сборник аннотаций XV совещания по ядерной спектроскопии и теории ядра. Дубна, 1978. | 2 р. 50 к. |
| Д3-11787 | Труды III Международной школы по нейтронной физике. Алушта, 1978. | 3 р. 00 к. |
| Д13-11807 | Труды III Международного совещания по пропорциональным и дрейфовым камерам. Дубна, 1978. | 6 р. 00 к. |
| | Труды VI Всесоюзного совещания по ускорителям заряженных частиц. Дубна, 1978 /2 тома/ | 7 р. 40 к. |
| Д1,2-12036 | Труды V Международного семинара по проблемам физики высоких энергий. Дубна, 1978 | 5 р. 00 к. |
| Д1,2-12450 | Труды XII Международной школы молодых ученых по физике высоких энергий. Приморско, НРБ, 1978. | 3 р. 00 к. |
| | Труды VII Всесоюзного совещания по ускорителям заряженных частиц, Дубна, 1980 /2 тома/ | 8 р. 00 к. |
| Д11-80-13 | Труды рабочего совещания по системам и методам аналитических вычислений на ЭВМ и их применению в теоретической физике, Дубна, 1979 | 3 р. 50 к. |
| Д4-80-271 | Труды Международной конференции по проблемам нескольких тел в ядерной физике. Дубна, 1979. | 3 р. 00 к. |
| Д4-80-385 | Труды Международной школы по структуре ядра. Алушта, 1980. | 5 р. 00 к. |
| Д2-81-543 | Труды VI Международного совещания по проблемам квантовой теории поля. Алушта, 1981 | 2 р. 50 к. |
| Д10,11-81-622 | Труды Международного совещания по проблемам математического моделирования в ядерно-физических исследованиях. Дубна, 1980 | 2 р. 50 к. |
| Д1,2-81-728 | Труды VI Международного семинара по проблемам физики высоких энергий. Дубна, 1981. | 3 р. 60 к. |
| Д17-81-758 | Труды II Международного симпозиума по избранным проблемам статистической механики. Дубна, 1981. | 5 р. 40 к. |
| Д1,2-82-27 | Труды Международного симпозиума по поляризационным явлениям в физике высоких энергий. Дубна, 1981. | 3 р. 20 к. |
| Р18-82-117 | Труды IV совещания по использованию новых ядерно-физических методов для решения научно-технических и народнохозяйственных задач. Дубна, 1981. | 3 р. 80 к. |

Заказы на упомянутые книги могут быть направлены по адресу:
101000 Москва, Главпочтамт, п/я 79
Издательский отдел Объединенного института ядерных исследований

Ершова Т.А., Олейникова С.Г., Сычев П.П. Транспортабельная программа "TEXTA" для подготовки текстовой документации 10-82-463

Программа "TEXTA" предназначена для подготовки текстовой документации в книгоподобном виде. Написана на стандартном языке КОБОЛ. Это обеспечивает перенос /адаптацию/ программы на ЭВМ любого типа, имеющую компилятор с КОБОЛа. Описаны основные возможности программы, ее структура и некоторые алгоритмы. Приведены эксплуатационные характеристики программы для ЭВМ различного типа.

Работа выполнена в Лаборатории вычислительной техники и автоматизации. ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1982

Ershova T.A., Olejnikova S.G., Sychev P.P. TEXTA Portable Program for Preparation of Text Documentation 10-82-463

The TEXTA program is intended to produce text documentation in book-like form. It is written in COBOL standard. This provides a portability of program on any computer with COBOL compiler. Main options of program, its structure and some algorithms are described. Usage characteristics of program for some JINR computers are given.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1982

Перевод О.С.Виноградовой.