

СООБЩЕНИЯ
ОБЪЕДИНЕННОГО
ИНСТИТУТА
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ

ДУБНА



Ц 8400
М - 137

29/12-75

11 - 9008

Г.Л.Мазный

3766/2-75

УНИВЕРСАЛЬНЫЙ ПЕРЕКОДИРОВЩИК ПЕРФОКАРТ,
ПРОБИТЫХ ПОКОЛОННО

1975

11 - 9008

Г.Л.Мазный

УНИВЕРСАЛЬНЫЙ ПЕРЕКОДИРОВЩИК ПЕРФОКАРТ,
ПРОБИТЫХ ПОКОЛОННО

Объединенный институт
восточных исследований
БИБЛИОТЕКА

Мазный Г.Л.

11 - 9008

Универсальный перекодировщик перфокарт, пробитых по колонно

В работе описана созданная в ЛВТА ОИЯИ программа, предназначенная: 1. Для перекодировки перфокарт из одного (любого) по колонного кода в другой (любой) по колонный код. 2. Для замены символов в пределах одного кода. 3. Для выполнения комбинации указанных действий.

Схема работы программы оригинальна: символы перекодируются не в отдельности, а некоторыми нетривиальными группами.

Программа имеет небольшой объем, однако заслуживает отдельного сообщения в силу большой практической применимости (облегчает перевод программного материала с одной серии ЭВМ на другую серию ЭВМ, расширяет диапазон возможных устройств для первичной подготовки программного материала).

Работа выполнена в Лаборатории вычислительной техники и автоматизации, ОИЯИ.

Сообщение Объединенного института ядерных исследований
Дубна 1975

Появление универсальных алгоритмических языков программирования (ФОРТРАН /1,2/, АЛГОЛ /3,4/ и др.) создало предпосылки для съема одних и тех же программ на ЭВМ различных типов. Вместе с тем, имеются существенные различия в кодировке программ и данных на перфокартах, что зачастую вынуждает программиста при переходе на другую машину заново пробивать (и, следовательно, заново отлаживать) программный материал.

Так, например, фирмы CDC и IBM не смогли договориться о введении единой кодировки информации на перфокартах. Используемый системой математического обеспечения "ДУБНА" БЭСМ-6 в качестве одного из стандартных код CDC, дополненный русскими буквами, еще более отличается от кода КПК-12, принятого на машинах типа ЕС, который представляет собой код IBM, также дополненный русскими буквами, но в другой кодировке. Правда, введенные в мониторную систему "ДУБНА" управляющие карты /5/ ⁷IBM, ⁷CDC, ⁷IST позволяют вводить перфокарты, пробитые, в частности, в любой из перечисленных кодировок, однако выводится информация только в (расширенной) кодировке CDC. Не решен вопрос ввода информации, пробитой в других по колонных кодах (например, в коде устройств подготовки данных типа УПДК, предназначенных для машин типа "Минск", и т.п.). Переход на машины, "не понимающие" русских букв, вызывает проблему соответствующей замены этих букв в идентификаторах программ.

Предлагаемый перекодировщик предназначен:

1. Для перекодировки перфокарт из одного по колонного кода в другой по колонный код.
2. Для замены одних символов другими в пределах одного кода.
3. Для выполнения комбинации указанных действий.

Перекодировщик не выполняет никаких действий, связанных с другими носителями информации, так как, в частности, не перекодирует информацию ни в один из внутренних кодов мониторной системы.

Перекодировщик написан на автокоде "МАДЛЕН" и может ис-

пользоваться только на ЭВМ БЭСМ-6.

Таблица соответствий

Соответствие между исходными кодами символов и кодами символов для замены задаётся с помощью специальной таблицы соответствий. Таблица соответствий выполняется в виде двух последовательных перфокарт: эталон1 и эталон2. Эталон1 содержит исходные коды символов. Эталон2 содержит в соответствующих позициях коды символов для замены. Символы, коды которых должны остаться неизменными, задавать в таблице соответствий не обязательно. Это позволяет при ограниченной длине таблицы соответствий практически всегда достигать результата за один запуск (длина таблицы соответствий до 80 символов, на устройствах подготовки перфокарт обычно можно пробить до 84 символов, но алфавиты языков включают значительно меньшее их количество).

Примеры.

1. При перекодировке из кода CDC в код IBM эталон1 должен содержать отличающиеся символы в коде CDC, эталон2 - соответствующие символы в коде IBM.

2. При необходимости замены всех латинских букв "D" русскими буквами "Д" в пределах расширенной кодировки CDC таблица соответствий должна иметь вид:

эталон1 - D

эталон2 - Д

И тот, и другой символ пробивается в расширенном коде CDC в первой позиции карты.

3. При необходимости замены всех русских букв "У" латинскими буквами "Y" в пределах кодировки КПК-12 соответственно:

эталон1 - Y

эталон2 - У

Оба символа пробиваются в первых позициях карт в коде КПК-12.

4. Пусть у нас имеется задача, которая считается на БЭСМ-6, и пусть её необходимо посчитать на ЭВМ серии IBM. При перекодировке (если программа и данные выполнены в расширенном коде CDC) необходимо в эталон1 включить все символы, коды которых в CDC и IBM различны (естественно, в исходном коде) и русские буквы, отличающие-

ся по начертанию от латинских. Эталон2 должен быть пробит в коде IBM и содержать (в тех случаях, когда это возможно) в соответствующих позициях коды тех же символов. В позициях, соответствующих русским буквам, отличающимся от латинских, должны быть указаны символы для замены.

Перекодировка производится посимвольно, поэтому в качестве замены, скажем, для русской буквы "Ч" можно указать символ "С" или символ "4", но нельзя указать сочетание "СН".

Естественно, подобное "сжатие" алфавита может привести к случайному появлению совпадающих идентификаторов, поэтому такие замены должны тщательно продумываться, иначе программа потребует доработки.

Признак окончания массива

В качестве исходного может быть выбран любой код, поэтому выбор признака окончания массива предоставляется самому программисту. Это может быть перфокарта, пробитая самым произвольным образом (в частности, содержащая пробивки, отсутствующие в исходном коде). Естественно, она не должна совпадать ни с одной из содержащихся в перекодируемом массиве перфокарт.

Назовём признак окончания массива: эталон3.

Карта эталон3 должна присутствовать в пакете в двух экземплярах - непосредственно после карты эталон2 и в конце перекодируемого массива.

Ограничения

Карты эталон1, эталон2, эталон3 и карты перекодируемого массива не должны совпадать ни с одной из карт:

*BINARY, *END_BINARY, *END_FILE, *NAME.name

в коде CDC.

Они не должны также совпадать с "диспетчерским концом задачи" - перфокартой со всеми пробивками в первой и сорок первой колонках и только в них.

Карты, связанные с расположением перекодировщика

В простейшем случае это может быть стандартный массив самого перекодировщика. Перекодировщик может быть записан в личную или общую библиотеку, тогда это карты обращения к соответствующей библиотеке.

Возможно, в дальнейшем перекодировщик будет занесён в библиотеку служебных задач – тогда он будет вызываться приказом с операторского устройства и отпадёт необходимость в ряде дополнительных карт.

Имя головной программы перекодировщика: РТР.

Признак окончания массива при выводе

В качестве признака окончания массива при выводе перфорируется карта эталон3 в неизменном виде.

Структура пакета перфокарт при перекодировке

Пакет перфокарт для перекодировки формируется следующим

образом:

*NAME:name
*PASS:pass
*TIME:time
*ASSIGN PUNCH

Карты, связанные с расположением перекодировщика

*MAIN_PTP
*EXECUTE
*BINARY

эталон1

эталон2

эталон3

перекодируемый массив

эталон3

*END_BINARY

*END_FILE

"диспетчерский конец задачи"

Длина перекодируемого массива

Перекодируемый массив должен состоять не менее, чем из одной перфокарты. Ограничений сверху на длину перекодируемого массива перекодировщик не накладывает и они определяются задачей ввода мониторной системы (примерно до 2.5 тысяч перфокарт).

Изложенным исчерпывается постановка задачи и, по существу, инструкция по использованию перекодировщика. Перейдём теперь к описанию принципа его работы.

Используемые системные подпрограммы

Вся работа со вводом и выводом в перекодировщике сведена к работе только с образами перфокарт. Образ перфокарты – это массив в памяти машины длиной 24 слова. Младшие 10 разрядов каждого слова массива содержат следующую информацию:

- 1 слово – левую половину 1 строки перфокарты,
- 2 слово – правую половину 1 строки перфокарты,
- 3 слово – левую половину 2 строки перфокарты,
- 4 слово – правую половину 2 строки перфокарты и так далее.

Именно таким образом представляется информация для ввода и вывода перфокарт с помощью экстракодов диспетчера БЭСМ-6 /7,8/.

Перекодировщиком используются следующие подпрограммы мониторной системы /5/:

- BINARY – ввод образа очередной перфокарты,
- PUNON – установка в начальное состояние буфера для перфорации,
- PUNCH – занесение образа перфокарты в буфер для перфорации (буфер перфорируется, в частности, по заполнению),
- PUNOFF – перфорация незаполненного полностью буфера,
- STOP* – подпрограмма завершения счёта задачи.

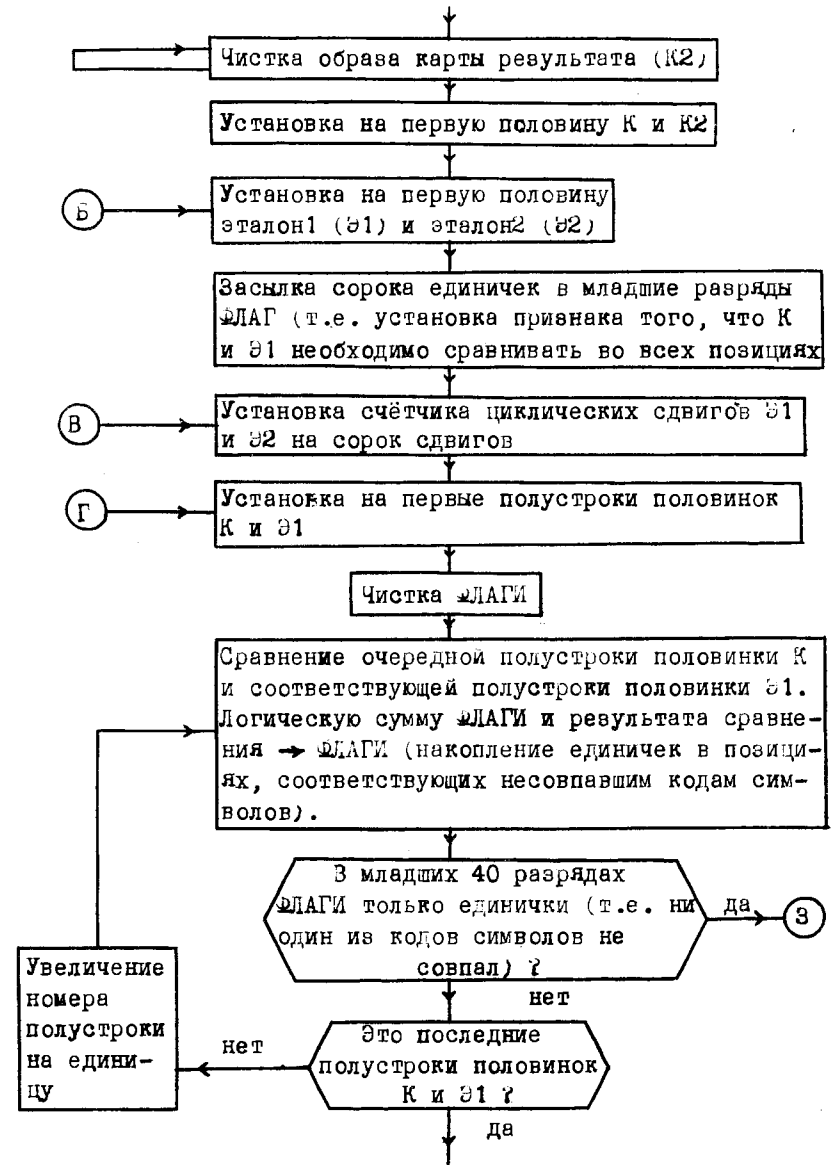
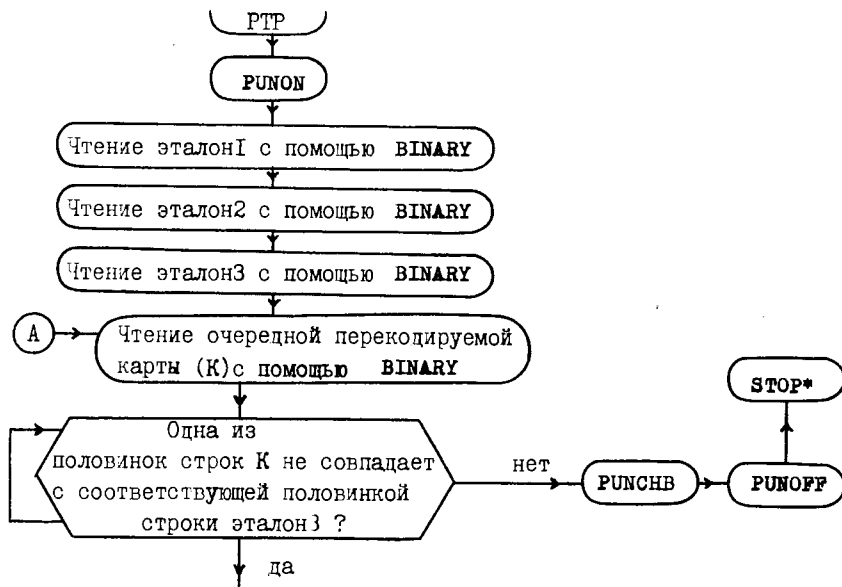
Из сказанного вытекает, что формат образа карты устроен самым неудобным образом для перекодировки перфокарт с построчной пробивкой. Обычно (например, в соответствующих подпрограммах мониторной системы "ДУБНА") в этом случае прибегают к специальным приёмам "переворота карты" и "обратного переворота карты", причём каждый перекодируемый символ рассматривается и обрабатывается в отдельности.

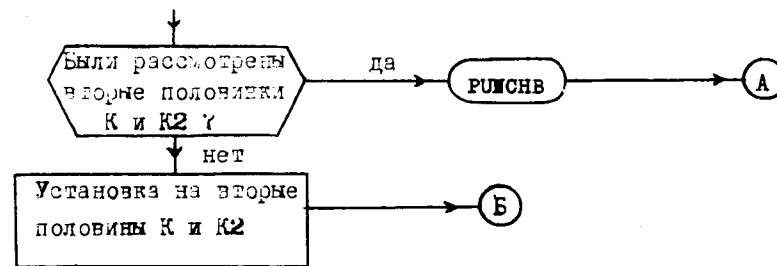
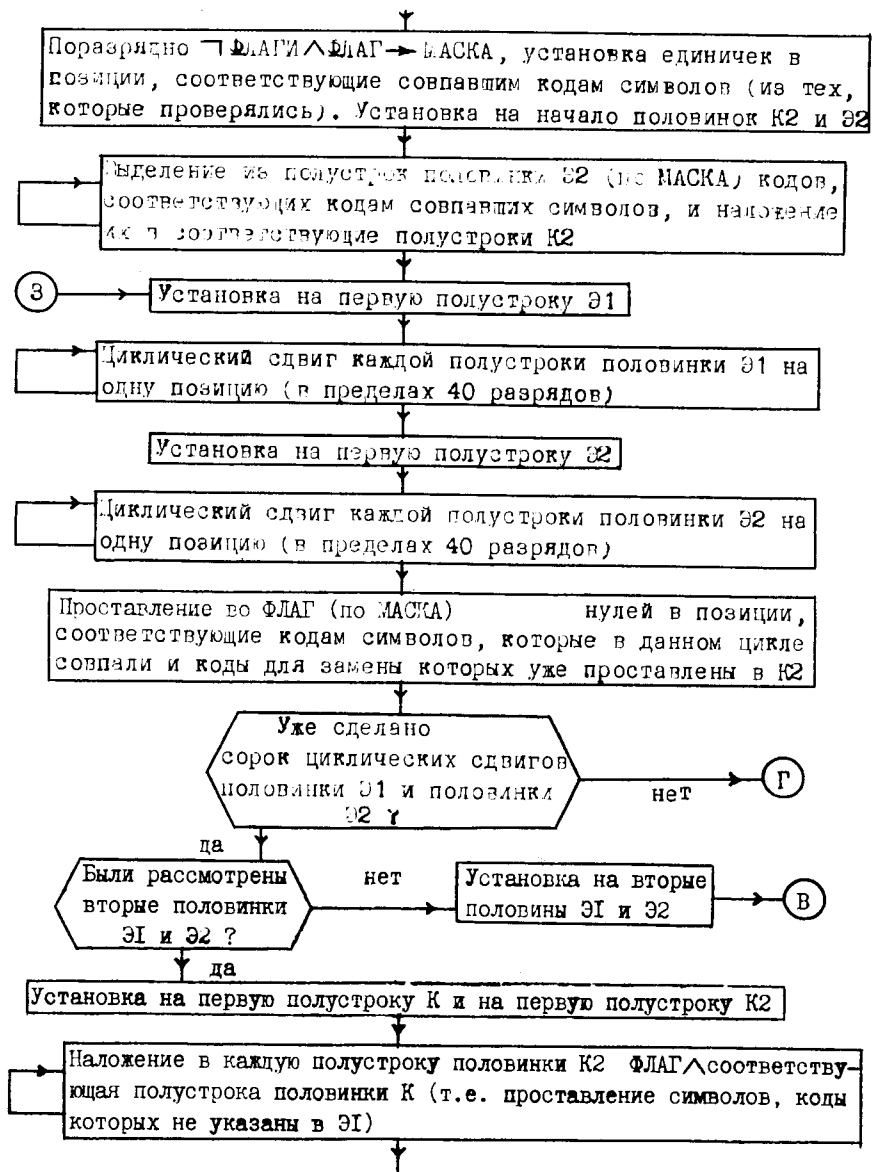
Схема работы перекодировщика

Предлагаемый перекодировщик лишён указанных недостатков. Схема работы перекодировщика оригинальна в том смысле, что:

1. Образ карты не подвергается "перевороту" и, следовательно, "обратному перевороту".
2. Символы не рассматриваются в отдельности, а перекодируются целыми группами, в результате чего экономится как счётное время, так и оперативная память. Вся работа ведётся "на пяточке" из пяти массивов длиной 24 слова каждый. Три из этих массивов статичны и содержат эталон1, эталон2 и эталон3, четвёртый содержит образ очередной перекодируемой карты, пятый - образ результата перекодировки. Все указанные карты "режутся" вертикально пополам и полученные части рассматриваются в отдельности. Поэтому в дальнейшем мы позволим себе пользоваться терминами "половинка карты" и "полкарты" для обозначения совокупности слов образа, имеющих только чётные или только нечётные номера.

Итак, схема работы перекодировщика:





При написании перекодировщика учитывалось, что циклы, содержащие не более трёх командных слов (пар команд) выполняются в машине существенно быстрее (все команды помещаются на быстрых регистрах, чем экономятся выборки из оперативной памяти).

Литература

1. Язык ФОРТРАН. Под редакцией В.П. Ширикова. ОИЯИ, II-4818, Дубна, 1969.
2. Г.И. Макаренко, А.В. Ракитский, А.И. Салтыков. ФОРТРАН. М. Знание, 1973.
3. Алгоритмический язык АЛГОЛ-60. Пересмотренное сообщение. Под ред. А.П. Ершова, С.С. Лаврова, М.Р. Шура-Бура. М., Мир, 1965.
4. Reidieter Bericht ueber die algorithmische Sprache ALGOL60, Akademie Verlag, Berlin, 1966.
5. Г.Л. Мазный. Мониторная система "ДУБНА". Руководство для пользователей. ОИЯИ, II-5974, Дубна, 1971.
6. А.И. Волков. Автокод MADLEN. ОИЯИ, II-5427, Дубна, 1970.
7. И.Н. Силин. Диспетчер ДД73 машины БЭСМ-6. ОИЯИ, ДЮ-7707, Дубна, 1974.
8. В.Ю. Веретеннов, М.И. Гуревич, А.В. Гусев, В.З. Житенёв, Н.С. Заикин, Л.Г. Каминский, О.Н. Ломидзе, И.Н. Силин, В.А. Федосеев, В.П. Шириков. Новый диспетчер для ЭВМ БЭСМ-6. ОИЯИ II-7059, Дубна, 1973.
9. Н.С. Заикин, О.Н. Ломидзе, Г.Л. Семашко, В.П. Шириков. Пакетная обработка на БЭСМ-6 в ОИЯИ. ОИЯИ, ДЮ-7707, Дубна, 1974.

Рукопись поступила в издательский отдел
24 июня 1975 года.