

Н. В. ВАРТАНЯН, Э. В. ЕГИАЗАРЯН, Р. Л. УРУТЯН

ОРГАНИЗАЦИЯ СЛОВАРЕЙ МАШИНЫ «ГАРНИ»

§ 1. Структура словарей

Реализация автоматического перевода (АП) связана с многократным обращением к словарям, при размещении которых в ЗУ машины возникают трудности, связанные с трудностями адресации, особенно, если словари велики. Кроме того, опыты по АП на универсальных электронных вычислительных машинах (ЭВМ) показывают, что значительная доля времени работы ЭВМ при переводе текста уходит на поиск слов в словаре. В связи с этим проблема организации словаря и поиск в словаре является одной из основных проблем математико-лингвистических исследований. Если условимся называть «вводимым словом» каждое отдельное слово переводимого текста, то под машинным словарем будем понимать автоматическое устройство, на вход которого поступает код вводимого слова, а на выходе получается некоторая информация, приписанная данному слову.

При одном из методов организации словарей [1] машинный словарь должен содержать таблицу констант, сопоставленных всем вводимым словам, и соответствующий алгоритм поиска. Такие словари характеризуются следующими признаками:

- 1) объемом памяти, который отводится под таблицу констант,
- 2) объемом памяти, который отводится под алгоритм поиска необходимой информации по вводимому слову;
- 3) сложностью переделки алгоритма или таблиц словаря при необходимости его пополнения или изменения;
- 4) средним временем поиска.

При указанном методе организации словарей возможен вариант, при котором словарь содержит все словоформы слов переводимых текстов. Очевидно, что общий объем памяти, используемый для хранения этих словарей, будет очень большим, хотя алгоритм поиска — сравнительно простым. Для экономии объема памяти машины используется метод замены словоформ набором основ, где под основой понимается наибольшая неизменная часть слова (которая, вообще говоря, может не совпасть с общепринятым представлением о грамматической основе слова).

В системе «Гарни» реализован последний метод организации словарей. Машинный словарь состоит из словарей: входного—русского и выходного—армянского языков, которые соответственно хранятся на двух магнитных барабанах.

Машинный словарь русского языка содержит последовательность словарных статей, которые состоят из основы, грамматической характеристики, а также некоторой вспомогательной информации, связанной с алгоритмом выборки. Словарные статьи расположены на дорожках МБ друг за другом без адресной локализации, причем из двух основ, одна из которых целиком вкладывается в другую, большая помещается перед меньшей. В русскую словарную статью входят следующие части (рис. 1).

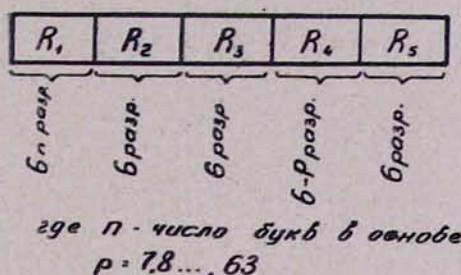


Рис. 1. Общий вид русской словарной статьи.

1. Основы (R_1) словарной статьи, которая закодирована побуквенно. На код одной буквы отводится одна секста (так мы называем набор из шести двоичных разрядов), причем старший разряд каждой сексты равен нулю.

2. Конец (R_2) основы, кодируемый одной секстой, старший разряд которой равен единице; оставшиеся пять разрядов используются для кодирования количества букв в данной основе (см. ниже).

3. Длина (R_3) характеристики, кодируемая одной секстой, указывающая на количество секст в характеристике.

4. Характеристика (R_4) к основе, количество разрядов которой кратно шести; она содержит грамматическую информацию ко всем словоформам слов, образующихся от данной основы.

5. Контрольный код (R_5), кодируемый секстой, получающейся при поразрядном сложении по *mod* 2 всех предшествующих секст данной статьи.

Из вышеуказанного следует, что длина основы словарных статей переменная; при этом в нашем словаре основа минимальной длины состоит из одной буквы, а максимальной—может содержать 31 букву. Что касается длины характеристики (R_4); то, исходя из временных требований (согласование работы МБ с УЦУ), минимальная длина R_4 составляет 7 секст, а максимальная длина определяется шестиразрядным числом R_3 , т. е. составляет 63 сексты.

Рассмотрим кодирование букв основы входного языка и знака R_2 . После объединения буквы e и $\bar{e}$, которое приводит к омонимии лишь в очень редких случаях, русский алфавит содержит 32 буквы, для кодирования которых достаточно пяти разрядов. Следовательно, возможно такое кодирование букв с помощью шести разрядов, при котором старший разряд каждой сексты равен нулю. Это помогает отличить код буквы от кода «конца основы» R_2 . При кодировании русских букв шестью разрядами мы исходим также из того, что алфавит выходного—армянского—языка состоит из 39 букв, для кодирования которых потребуется шесть двоичных разрядов. При применении такого кодирования русских букв из полного набора остаются 32 избыточных набора, которые используются следующим образом.

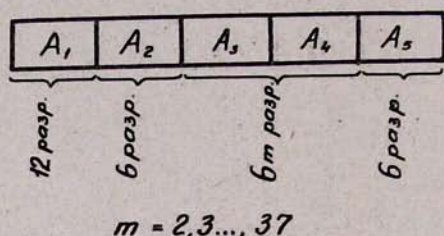


Рис. 2. Общий вид армянской словарной статьи.

Поиск в словаре входного языка производится на этапе морфологического анализа. Одной из основных функций морфологического анализа является членение каждого вводимого слова на основы и окончания. Принципиально возможны несколько вариантов членения. Вариант, который реализован на машине «Гарни», основан на подсчетах количества букв в основе, соответствующей данному вводимому слову. При этом, с целью экономии используемого оборудования, код, который показывает количество букв в основе, записывается на дорожке МБ, одновременно используя те 32 набора, которые были избыточны при кодировании букв основ. Так как длина основы не превышает 32 буквы, то пять двоичных разрядов достаточно для кодирования количества чисел этих букв. Таким образом для R_2 получим такой код, при котором в старшем разряде всегда имеется единица и при помощи которого определяется конец основы; а остальные пять разрядов—есть число, равное количеству букв, содержащихся в данной основе.

Машинный словарь выходного языка содержит последовательность армянских словарных статей, помещенных на дорожках барабана в порядке убывания их номеров. В армянскую словарную статью входят следующие части (рис. 2).

1. Номер словарной статьи (A_1), который кодируется двумя секстами; он является наименьшим номером среди номеров тех слов, которые получаются от данной основы.

2. Длина (A_2) информации, кодируемой одной секстой и указывающей на количество секст в (A_3) и (A_4).

3. Основа (A_3) словарной статьи, буквы которой закодированы секстами.

4. Характеристика (A_4), количество разрядов которой кратно шести; она содержит грамматическую информацию, необходимую для получения всех словоформ всех слов, образуемых от данной основы.

5. Контрольный код (A_5) аналогичен контрольному коду русской статьи.

§ 2. Алгоритм обращения к словарям

Как отмечалось выше, русская словарная статья содержит основу и информацию к ней. Они помещены на дорожках МБ непосредственно одна за другой без локализации адреса, по которому они записаны. Поиск требуемой статьи производится от начала дорожки по совпадению букв анализируемого слова с буквами каждой основы. Выбирается та первая статья, основа которой целиком вкладывается в данное слово. Выбранная основа будет наибольшей по длине, так как основы расположены в порядке убывания их длины. Основной операцией при реализации поиска является сравнение букв вводимого слова и основы. Хранение очередной буквы основы во время сравнения осуществляется на шестиразрядном регистре (РБ). Само сравнение осуществляется с помощью шестиразрядного дешифратора (ДС), на входы которого поступают выходы из регистра РБ и регистра связи (РС), на который с регистра P_3 ОУ подается очередная буква анализируемого слова. Дешифратор ДС выдает сигнал только при несовпадении. Выходы регистра РБ поступают также на вход дешифратора конца основы (ДК), который одновременно с сравнением фиксирует конец основы. После нахождения основы с дорожки МБ в ДОЗУ выдается вся оставшаяся часть статьи (начиная с кода A_2 до конца). Во время поиска статьи идет проверка правильности приема информации с дорожки МБ. Для этого выходы регистра РБ поступают в регистр контроля ($P_{\text{контр.}}$). В конце словарной статьи записан контрольный код (R_5, A_5), который сравнивается с суммой, полученной на $P_{\text{контр.}}$ путем сложения по $\text{mod } 2$ всех секст, входящих в данную статью (кроме кода R_5, A_5). При несовпадении кодов устройство выдает сигнал, который может быть использован для программной организации повторного поиска с начала той же дорожки. Определение конца словарной статьи осуществляется следующим образом: код длины характеристики, записанный непосредственно после признака конца основы в русской словарной статье или после номера в армянской словарной статье, передается в счетчик длины СчД. После поступления каждой сексты информации из содержимого СчД вычитается единица. Появление кода «0» в счетчике означает конец данной статьи. Выбор статьи осуществляется с помощью АП, ко-

торый управляет работой дешифратора ДС; разрешение сравнения выдается каждый раз с начала статьи. Сравнение прекращается после поступления сигнала несовпадения или конца основы (или номера—для армянского словаря).

Опишем метод поиска армянских словарных статей, осуществляемый аналогично поиску русских статей, с той лишь разницей, что здесь в словаре должно быть найдено то слово, номер α_0 которого выработан в результате русского анализа и армянского синтеза. В армянском словаре, так же как и в русском, слова, имеющие общую основу, сведены в одну статью, номер α которой является наименьшим из номеров слов, получающихся от данной основы. Следовательно, из словаря должна быть выбрана та статья, номер которой является наибольшим из номеров тех статей, которые не превосходят номера переводящего слова. Сам поиск осуществляется сравнением номера α_0 с номерами статей на данной дорожке (выбор дорожки осуществляется программно). Так как словарные статьи на дорожке расположены в порядке убывания их номеров, то искомой статьей является та первая статья, для которой имеет место неравенство $\alpha_0 \geq \alpha$. Если это условие обозначить χ_1 , а противоположное условие, $\alpha_0 < \alpha$, обозначить через χ_2 и построить автомат поиска, аналогичный вышеуказанному, то легко видеть, что автоматы, реализующие поиск в обоих словарях, являются изоморфными. Заметим, что при поиске армянских статей в качестве входного сигнала χ_2 берется признак конца номера статьи.

Для реализации описанного метода выборки статей из русского и армянского словарей в машине «Гарни» имеется так называемое устройство ассоциативного поиска статей (УАПС). Дадим краткое описание его работы (рис. 3). Перед каждым поиском программно определяется и выдается на регистр РИ₂ ОУ номер дорожки, в которой хранится искомая словарная статья. С помощью синхронизирующего сигнала «0» МБ процесс сравнения начинается с начала дорожки, независимо от того, в какой момент поступит разрешение на поиск.

При этом, если производится поиск русских словарных статей, на регистр Р₃ выдаются первые пять букв анализируемого слова, после поступления первой буквы основы на регистр РБ, на регистр РС принимается старшая секста регистра Р₃ и выдается разрешение на их сравнение. Если в пределах первых пяти букв сигнал несовпадения не вырабатывается, то после проверки пятой буквы на регистр Р₃ выдается содержимое следующего отрезка той линии ДОЗУ, в которой записано анализируемое слово, и сравнение продолжается с оставшимися буквами основы. Этот процесс продолжается до появления либо сигнала «несовпадение», либо кода R_2 на РБ. Если до появления на регистре РБ кода R_2 произошло совпадение, то грамматическая характеристика (начиная с R_2) через регистр Р₄ записывается в определенную линию ДОЗУ. При появлении же сигнала «несовпадение» на регистре РЗУ восстанавливаются начальные пять букв анализируемого слова, данная статья пропускается, и весь процесс повторяется со следующей ста-

тей. Если до конца дорожки искомая статья не обнаруживается, то выдается сигнал об отсутствии данной статьи на данной дорожке.

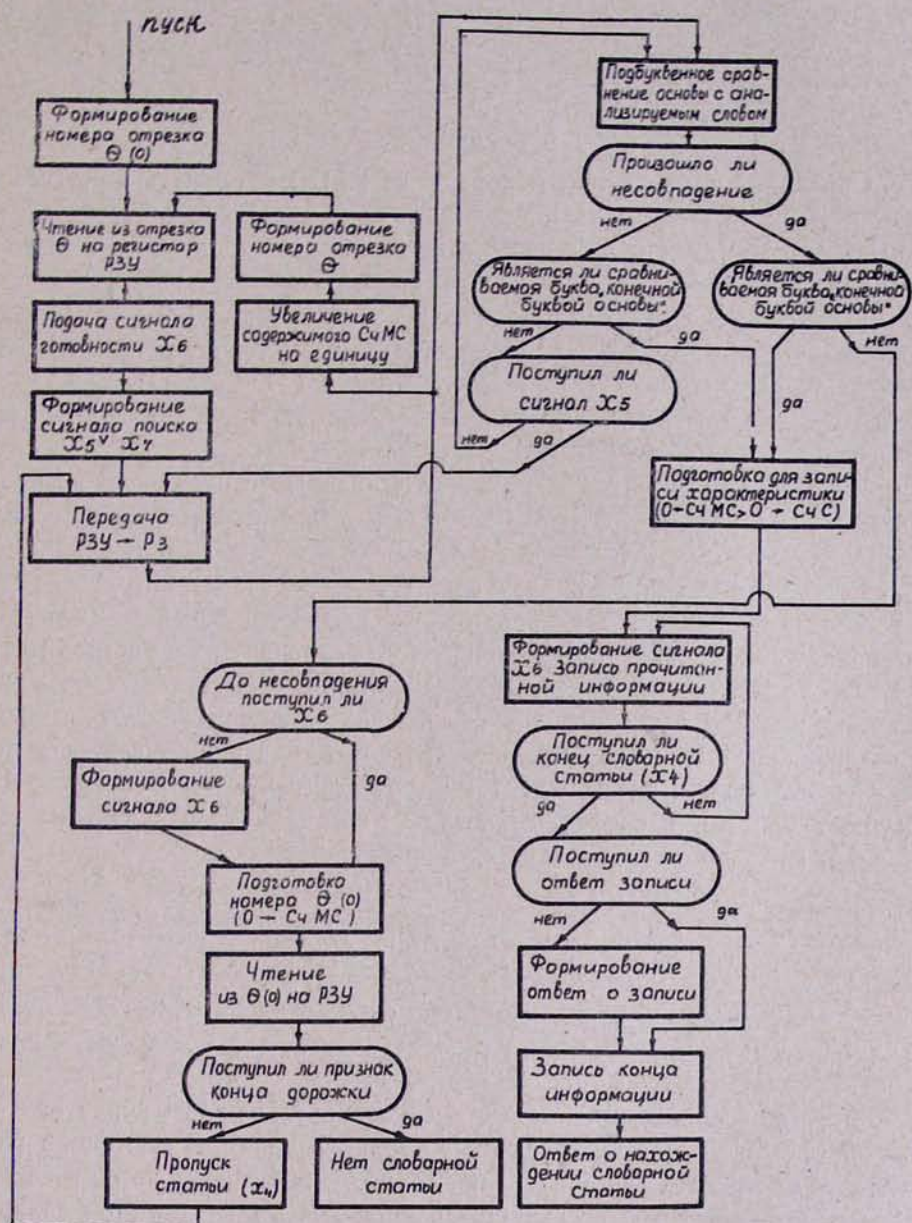


Рис. 3. Алгоритм поиска словарных статей.

При поиске армянской словарной статьи на старшие двенадцать разрядов регистра P_3 предварительно выдан номер армянской словарной статьи. Первые две сексты каждой статьи с дорожки МБ через регистры РВ и РС выдаются в левые 12 разрядов регистра P_4 , после чего производится вычитание содержимого регистра P_3 из содержимого P_4 .

Если $\tilde{P}_4$ меньше $\tilde{P}_3$, то эта статья пропускается и на P_4 выдается номер следующей статьи. Если оказывается, что $\tilde{P}_4 > \tilde{P}_3$, то это означает, что данная статья является искомой, и информация к ней через P_4 записывается в определенную линию ДОЗУ.

§ 3. Абстрактный синтез УАПС

Описанный в предыдущем параграфе УАПС машины «Гарни» состоит из двух МБ для хранения русского и армянского словарей, автомата поиска (см. ниже) и нижеследующих узлов проверки указанных там условий.

1) Узел сравнения вводимого слова с основой. Этот узел формирует выходной сигнал γ_1 , если анализируемые буквы равны. В противном случае узел выдает сигнал γ_2 .

2) Узел фиксации признака конца основы, при поступлении которой на регистр РБ, формируется сигнал γ_3 .

3) Узел фиксации конца словарной статьи. Этот узел при помощи кодов R_2 , A_2 и счетчика СчД определяет конец словарной статьи, выдавая сигнал γ_4 .

Рассмотрим принцип работы каждого узла в отдельности.

Узел сравнения (УС) представляет шестиразрядный дешифратор несовпадения кодов, которые подаются на его входы из регистров РБ и РС. Поскольку сравнение, прием, выдача производятся секстами, а информация с МБ поступает последовательно, возникает необходимость формирования синхронизирующего сигнала 6М, что достигается с помощью счетчика с коэффициентом пересчета 6, на вход которого поступают маркерные синхронизирующие импульсы. Узел сравнения выполняет следующую логическую функцию:

$$(\bar{p}_1 \wedge \bar{\beta}_1) \vee (\bar{p}_1 \wedge \beta_1) \vee (\bar{p}_2 \wedge \bar{\beta}_2) \vee (\bar{p}_2 \wedge \beta_2) \vee \dots \vee (\bar{p}_6 \wedge \bar{\beta}_6) \vee (\bar{p}_6 \wedge \beta_6),$$

где $p_1, p_2, \dots, p_6$ — разряды буквы основы, а $\beta_1, \beta_2, \dots, \beta_6$ — разряды букв анализируемого слова. При этом если логическая функция равна 1, то выдается сигнал γ_1 , а если равна 0 — γ_2 . Блок-схема узла сравнения представлена на рис. 4.

Узел фиксации признака конца основы. Так как код конца основы отличается от кодов буквы основы наличием единицы в старшем разряде, то для обнаружения признака конца основы достаточно в регистре РБ анализировать только старший разряд каждой буквы. При обнаружении признака конца основы выдается сигнал γ_3 , и сравнение прекращается.

Узел фиксации признака конца статьи. Как видно из рис. 1 и 2, после кода конца основы или номера записан шестиразрядный код, указывающий длину характеристики. Поэтому для подсчета количества секст в характеристике в состав УАПС включен шестиразрядный счетчик СчД (рис. 5). Работа счетчика управляется автоматом с двумя

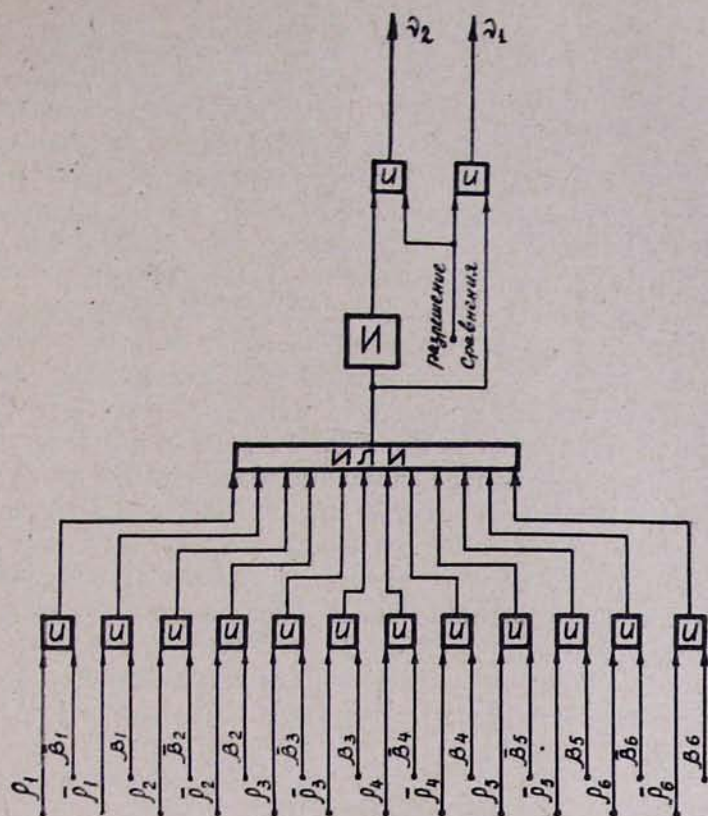


Рис. 4. Узел сравнения.

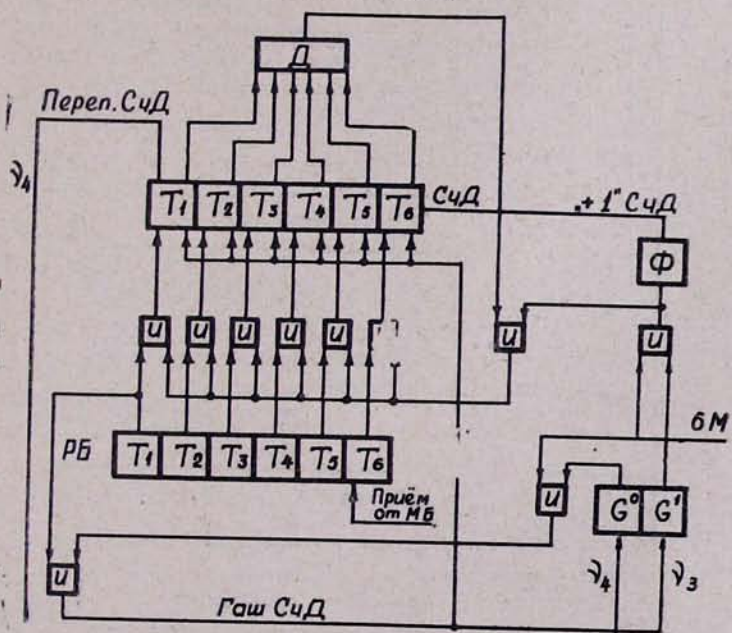


Рис. 5. Счетчик длины характеристики.

устойчивыми состояниями (G^0 и G^1) и входным алфавитом v_3 и v_4 . Входные клапаны счетчика блокированы либо состоянием автомата (G^0), либо содержимым самого счетчика (если содержимое счетчика отлично от нуля). С приходом сигнала v_3 автомат переходит в состояние G^1 , а счетчик гасится. Открываются входные клапаны счетчика, и из регистра РБ следующая секста характеристики R_3 в дополнительном коде принимается в счетчик. Затем закрываются входные клапаны, но остается открытым вход пересчета счетчика. Далее каждый поступающий сигнал 6М увеличивает содержимое счетчика на единицу. Очевидно, что признак переполнения счетчика СЧД будет означать конец статьи. С приходом сигнала v_4 автомат возвращается в состояние G^0 .

Во всех операциях, связанных со словарем, в качестве буферных регистров используются регистры РИ₄, Р₃ и Р₄ операционного узла.

Координацию работы трех описанных узлов и МБ, с одной стороны, с работой УЦУ машины «Гарни» — с другой, осуществляет специальный автомат, названный нами автоматом поиска (А), к синтезу которого мы переходим. На входы этого автомата должны быть подключены выходы узлов сравнения, фиксации конца статьи, т. е. на входы автомата А должны поступать сигналы v_1 , v_2 , v_3 , v_4 . Для составления входного алфавита автомата рассмотрим все комбинации этих сигналов (табл. 1). Анализ таблицы показывает:

Таблица 1
Входной алфавит автомата

№ №	v_4	v_3	v_2	v_1	Выполняемая операция.	
1	0	0	0	0	Не рассматривается.	
2	0	0	0	1	Продолжается сравнение слова с основой.	
3	0	0	1	0	Сравнение прекращается. Пропускается оставшая часть словарной статьи.	
4	0	0	1	1	Не имеет места.	
5	0	1	0	0	Не имеет места.	
6	0	1	0	1	Сравнение прекращается.	Если до поступления v_3 произошло совпадение, выдается характеристика статьи.
7	0	1	1	0		Если до поступления v_3 произошло несовпадение, то имеет место пункт 3.
8	0	1	1	1	Не имеет места.	
9	1	0	0	0	Конец словарной статьи.	
10-16						Не имеют места.

а) одновременно на входах узла сравнения не могут появляться сигналы γ_1 и γ_2 ;

б) так как осуществляется упорядоченность формирования сигналов $\gamma_1, \gamma_2, \gamma_3, \gamma_4$, при поиске в словарях будут иметь место лишь те комбинации этих сигналов, которые помещены в строках 1, 2, 3, 6, 7, 9 таблицы. Обозначим случай 2 через x_1 , 3 — x_2 , 6 или 7 — x_3 , 9 — x_4 . Символы x_1, x_2, x_3, x_4 составляют входной алфавит автомата (случай 1 не рассматривается). Заметим, что при составлении табл. 1 реальные длины регистров и временные характеристики машины не учитывались. Предполагалось, что длина регистров достаточна для записи вводимого слова или информационной части максимальной длины; при этом сдвиг, прием или выдача содержимого этих регистров производится без потери времени. Задача учета реальных длин регистров, временных характеристик УЦУ и частоты следования синхронизирующих импульсов МБ решается путем согласования времени одного цикла УЦУ (считывание и запись) со временем, которое необходимо для прохождения 30 маркерных синхронизирующих импульсов (длина регистров при работе со словарем — 30 разрядов) (см. ниже). Тогда на вход автомата необходимо подать следующие дополнительные сигналы:

x_5 — сигнал о прохождении 30 маркерных синхронизирующих импульсов МБ; этот сигнал формируется с помощью счетчика секст (СчС), на вход которого поступают сигналы 6М; СчС одновременно выполняет роль переключателя, который обеспечивает последовательную выдачу секст P_3 для сравнения или последовательное заполнение P_4 характеристикой статьи;

x_6 — сигнал готовности УЦУ для очередного запуска (этим сигналом является сигнал «ответ ЗУ» после выборки из отрезка или записи на отрезке ДОЗУ);

x_7 — синхронизирующий импульс «0» МБ; этот сигнал обеспечивает поиск с начала дорожки, независимо от момента поступления сигнала разрешения поиска.

Синтезируемый автомат должен формировать следующие выходные сигналы:

y_1 — сигнал подготовки очередных пяти букв вводимого слова;

y_2 — сигнал подготовки первых пяти букв вводимого слова после очередного события несовпадения;

y_3 — сигнал подготовки номера начального отрезка определенной линии ДОЗУ для записи прочитанной информации;

y_4 — сигнал записи очередных 30 разрядов прочитанной информации;

y_5 — сигнал о завершении записи прочитанной информации (т. е. сигнал о нахождении словарной статьи).

Поиск одной словарной статьи делится на два этапа:

а) выбор из ДОЗУ последовательных отрезков вводимого слова (или номера переводящего слова) и сравнение этих отрезков с отрезками, поступающими с МБ,

б) пропуск статьи (при несовпадении) или выдача для записи в ДОЗУ информационной части статьи.

Итак, событие, представленное в автомате сигналом y_5 , записывается в виде:

$$S = x_6 x_7 \{ \{ x_6 x_5 \} (x_2 x_3 x_6 \vee x_2 x_6 x_3 \vee x_6 x_2 x_3) x_6 x_4 \} \{ x_6 x_5 \} (x_3 x_6 \vee \vee x_6 x_3) x_5 \{ x_6 x_5 \} (x_4 x_6 \vee x_6 x_4) x_6. \quad (1)$$

Для строгого определения структуры слов, входящих в событие, необходимо было бы учесть поступление сигнала x_1 . Однако это привело бы к резкому осложнению записи. Между тем очевидно, что если после сигнала x_7 или очередного сигнала x_5 не появились сигналы x_2 или x_3 , то это означает, что в течение всего времени сравнения происходило совпадение. Напомним, что подсчет длины поступающих слов производится с помощью счетчика СЧС.

Одновременно обратим внимание на тот факт, что сигналы x_1 и x_2 являются запрещенными после обнаружения несовпадения или конца основы. Поэтому после появления этих сигналов (x_2 или x_3) должен выключаться узел сравнения и заново включаться после появления сигнала x_4 .

При рассмотрении регулярного выражения (1) выясняется, что наличие автомата G допускает дальнейшее упрощение этого выражения.

Действительно, если из (1) выделим слово $p = x_2 x_3 x_6 \vee x_2 x_6 x_3 \vee x_6 x_2 x_3$, то нетрудно убедиться, что входным сигналом x_3 фиксируется только конец проверяемой основы, который в свою очередь позволяет определить конец данной словарной статьи. Поскольку определение конца словарной статьи, как показано выше, производится с помощью автомата G (совместно с счетчиком СЧД), то входной сигнал x_3 в слове p можно опустить.

Тогда для p получим

$$p = x_2 x_6 \vee x_6 x_2.$$

Подставив значение p в выражение (1), получим:

$$S = x_6 x_7 \{ \{ x_6 x_5 \} (x_2 x_6 \vee x_6 x_2) x_6 x_4 \} \{ x_6 x_5 \} (x_3 x_6 \vee \vee x_6 x_3) x_5 \{ x_6 x_5 \} (x_4 x_6 \vee x_6 x_4) x_6. \quad (2)$$

Хотя регулярное выражение (2) полностью описывает работу автомата поиска, для формирования выходных сигналов $y_1, \dots, y_4$ необходимо составить соответствующие регулярные выражения, после наступления которых формируются выходные сигналы $y_1, \dots, y_4$. Оказывается, если взять разные начальные отрезки слов включенными в событие (2), то легко сформировать любой из сигналов $y_1, \dots, y_4$.

Так, например.

Формирование сигнала y_1 . Сигналом y_1 производится прием на регистр P_3 из регистра ЗУ очередных пяти букв анализируемого

слова и подготовка следующей пятерки на регистр ЗУ. Вначале это производится с помощью сигнала x_7 , а в остальных случаях — сигналом x_5 .

$$S_1 = x_6 x_7 [x_6 x_5] |_{y_1}.$$

Формирование сигнала y_2 . Сигналом y_2 осуществляется подготовка первых пяти букв вводимого слова после очередного события несовпадения (вначале поиска это производится командой, что здесь не учитывается)

$$S_2 = x_6 x_7 [x_6 x_5] (x_2 x_6 \vee x_6 x_2) |_{y_2}.$$

Член $x_2 x_6 \vee x_6 x_2$ выражения S_2 обуславливается временными характеристиками УЦУ и синхронизирующими сигналами МБ (рис. 6).

Формирование сигнала y_3 . Сигналом y_3 производится подготовка номера отрезка, в который записываются начальные тридцать разрядов информационной части статьи:

$$S_3 = x_6 x_7 [|x_6 x_5] (x_2 x_6 \vee x_6 x_2) x_6 x_4 | x_6 x_5] (x_3 x_6 \vee x_6 x_3) |_{y_3}.$$

Формирование сигнала y_4 . С помощью сигнала y_4 формируется очередной номер отрезка ДОЗУ, в который записывается продолжение информации, прочитанной из МБ:

$$S_4 = x_6 x_7 [|x_6 x_5] (x_2 x_6 \vee x_6 x_2) x_6 x_4 | x_6 x_5] (x_3 x_6 \vee x_6 x_3) x_5 [x_6 x_5] (e \vee x_4 x_6 \vee x_6 x_4) |_{y_4}.$$

Формирование сигнала y_5 . Этим сигналом обуславливается конец поиска словарной статьи, и машина переходит к выполнению следующей команды:

$$S_5 = x_6 x_7 [|x_6 x_5] (x_2 x_6 \vee x_6 x_2) x_6 x_4 | x_6 x_5] (x_3 x_6 \vee x_6 x_3) x_5 [x_6 x_5] (x_4 x_6 \vee x_6 x_4) x_6 |_{y_5}.$$

Очевидно, что $S_5 = S$. Таким образом, для построения автомата достаточно рассмотреть только выражение (2).

После составления отмеченной таблицы переходов и объединения колонок, имеющих одинаковые внутренние состояния и отмеченных одинаковыми выходными сигналами, получим упрощенную таблицу переходов автомата P (табл. 2). Граф автомата P представлен на рис. 7.

Проанализируем выражение S , разделив его на три части:

$$S = S' \cdot S''' \cdot S'',$$

где

$$S' = x_6 x_7 [|x_6 x_5] (x_2 x_6 \vee x_6 x_2) x_6 x_4 |$$

$$S'' = x_5 [x_6 x_5] (x_4 x_6 \vee x_6 x_4) x_6;$$

$$S''' = [x_6 x_5] (x_3 x_6 \vee x_6 x_3).$$

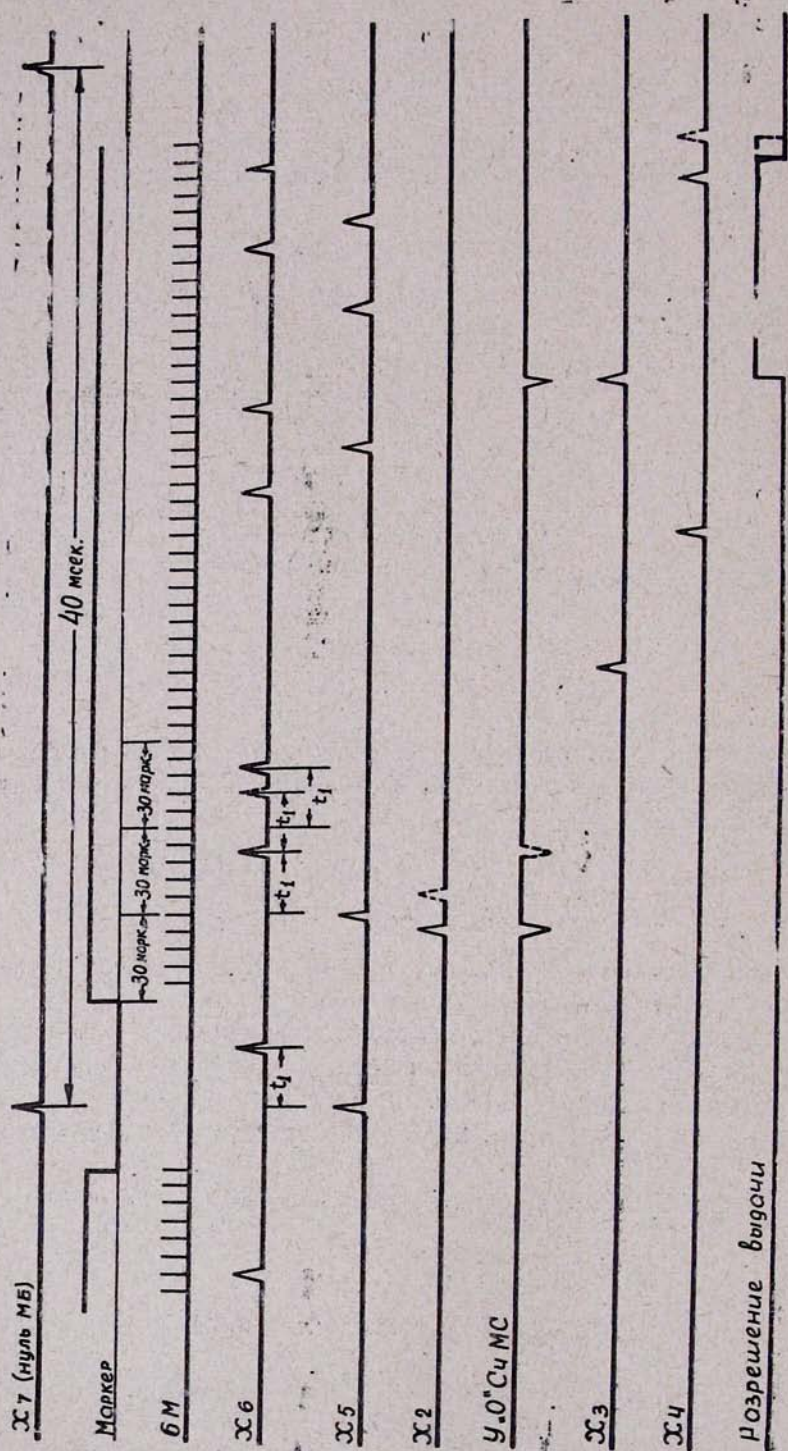


Рис. 6. Временная диаграмма поиска словарных статей.

Таблица 2

Упрощенная таблица переходов автомата Р

выходной сигнал	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}
состояние входной сигнал	0	1	2	3	4	5	6	7	8	9	10	11	12	13
x_2			5	4										
x_3			7	8										
x_4							2			12	11			
x_5				2					9		9			
x_6	1		3		6	4		8		10		13	11	
x_7		2												

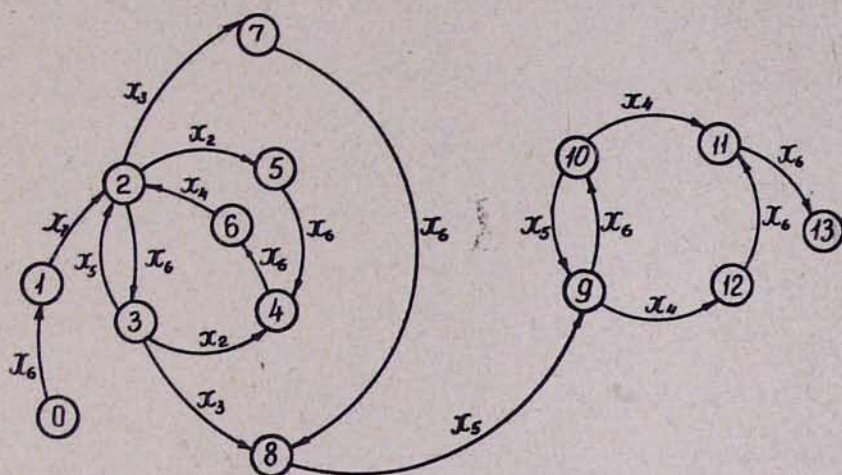


Рис. 7. Граф-схема автомата Р.

Рассмотрим выражение S' .

Входной сигнал x_4 служит для определения начала каждой статьи при обнаружении несовпадения в предыдущем сравнении. Для определения начала первой статьи на отдельной дорожке МБ записывается управляющий сигнал x_7 («0» МБ), означающий начало дорожки. Хотя сигнал x_7 используется в других целях устройства, относительно автомата сигналы x_4 и x_7 выполняют одинаковые функции.

Объединяя сигнал x_7 с сигналом x_4 , для выражения S' получим

$$S' = x_6 x_4 \{ [x_6 x_5] (x_2 x_6 \vee x_6 x_2) x_6 x_4 \}.$$

Теперь построим автоматы Q' и Q'' (рис. 8, 9), представляющие события S' и S'' . Эти автоматы обладают следующими свойствами:

- 1) они изоморфны;
- 2) автомат Q' работает только при сравнении букв вводимого слова с буквами основы либо номера переводимого слова с номером статьи;

3) автомат Q'' работает только после нахождения искомой статьи и обеспечивает запись характеристики основы в ДОЗУ.

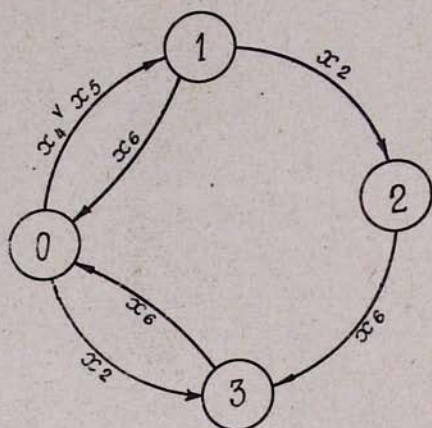


Рис. 8. Граф-схема автомата Q' .

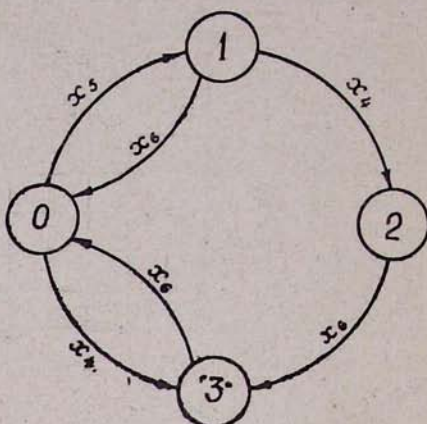


Рис. 9. Граф-схема автомата Q'' .

Если, следовательно, построить некоторый автомат, играющий роль переключателя, то автоматы Q' и Q'' можно объединить в одном Q . Для этого, пользуясь табл. 1, составим регулярное выражение

$$S_6 = \{x_2x_3x_4\}x_3x_4,$$

представленное автоматом U , граф-схема которого изображена на рис. 10.

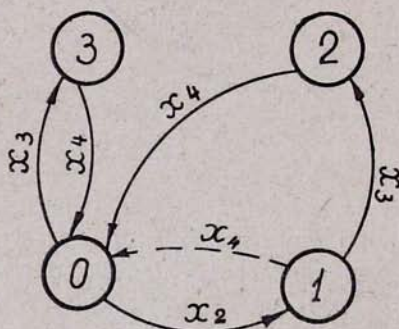


Рис. 10. Граф-схема автомата U .

Можно проверить, что если в этом автомате в качестве заключительных выбрать состояния 3 и 1, 2, то он будет играть роль указанного переключателя. Если же в качестве заключительных выбрать состояния 2 и 3, то этот же автомат будет выполнять функции автомата G , определяющего конец словарной статьи. Выражение S''' реализуется состояниями 0, 3 автомата U и состояниями 0, 1 автомата Q . Итак, нами построены два автомата Q и U , выполняющие роль и автомата P и автомата G .

Граф системы автоматов Q и U представлен на рис. 11. Из этого графа видно, что из реально возможных шестнадцати состояний в системе используются одиннадцать состояний. Между тем автоматы P и G требуют для своей реализации пять элементарных автоматов, а автомат, полученный с помощью регулярного выражения (1), включает в себя шестнадцать состояний.

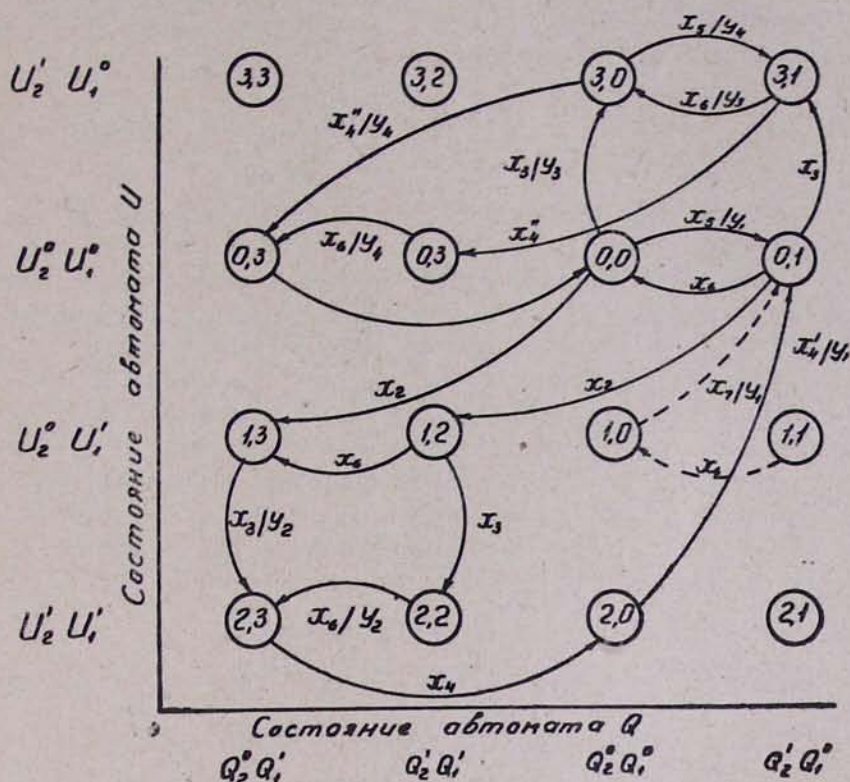


Рис. 11. Граф системы автоматов Q и U .

Одновременно очевидна простота этапа структурного синтеза системы автоматов Q и U .

Однако следует отметить, что объединение входных сигналов x_4 и x_7 вызывает некоторые технические трудности при обеспечении подачи информации с начала дорожки независимо от времени поступления сигнала поиска (x_0). Для решения этой задачи используются состояния $U_2^0 U_1^1 Q_2^1 Q_1^0$ и $U_2^0 U_1^1 Q_2^0 Q_1^0$ (ребра, идущие из этих состояний (рис. 11), проведены пунктирной линией). А для системы автоматов Q и U это означает, что в качестве начальных состояний следует рассматривать состояния 1 (вместо состояний нуля).

§ 4. Структурный синтез УАПС

Мы опишем структурный синтез автомата Q , который выполнен с помощью двух элементарных автоматов Q_1 и Q_2 . Кодирование внут-

ренных состояний автомата Q осуществляется в несколько этапов, описанных в [7, 8]. (Структурный синтез автомата U производится аналогичным образом).

На первом этапе кодирования выписываются все допустимые r -разделения множества внутренних состояний. Для автомата с четырьмя внутренними состояниями существуют три допустимых двоичных разделения:

$$r_1 = \{\overline{01}, \overline{23}\},$$

$$r_2 = \{\overline{02}, \overline{13}\},$$

$$r_3 = \{\overline{03}, \overline{12}\}.$$

На втором этапе при помощи таблицы переходов (табл. 3) должны быть выписаны все π -разделения, соответствующие r -разделению. Однако, как видно из табл. 3, синтезированный нами автомат является частичным, в силу чего мы применяем следующий метод. В пустые клетки табл. 3 вписываются цифры, соответствующие номеру столбца, записанному в верхней строке этой таблицы (табл. 4). При этом методе таблица выходов автомата остается неизменной. Теперь по составленной табл. 4 записываются все π -разделения. Если на этом этапе удастся получить оптимальное кодирование внутренних состояний автомата, то на этом кодирование прекращается. В противном случае мы вновь возвращаемся к методу работы [8]. В нашем случае переход от r -разделений к π -разделениям дает:

$$\pi_1 \rightarrow r_1 = \{\overline{01}, \overline{23}\} \rightarrow \{\overline{01}, \overline{2}, \overline{3}\},$$

$$\pi_2 \rightarrow r_2 = \{\overline{02}, \overline{13}\} \rightarrow \{\overline{03}, \overline{1}, \overline{2}\},$$

$$\pi_3 \rightarrow r_3 = \{\overline{03}, \overline{12}\} \rightarrow \{\overline{0}, \overline{12}, \overline{3}\}.$$

Таблица 3

Таблица переходов автомата Q

Входы \ Состояние	0	1	2	3
X_2	3	2	—	—
X_4	3	2	—	—
X_5	1	—	—	—
X_6	—	0	3	0

Таблица 4

Входы \ Состояние	0	1	2	3
X_2	3	2	2	3
X_4	3	2	2	3
X_5	1	1	2	3
X_6	0	0	3	0

Если начальному состоянию автомата присвоить код 00, то после взятия любой пары из π -разделения получим:

$$\text{по переменной } Q_1: \overline{01} \rightarrow 0; \overline{23} \rightarrow 1;$$

$$\text{по переменной } Q_2: \overline{03} \rightarrow 0; \overline{12} \rightarrow 1.$$

С помощью табл. 3 построим кодированную таблицу автомата (табл. 5). Кодированная таблица автомата U приведена на табл. 6.

Таблица 5

Кодированная таблица автомата Q

Входы Q_2, Q_1	00	01	10	11
x_2	01	—	11	—
x_4	01	—	11	—
$x_4' \vee x_5$	10	—	—	—
x_6	—	00	00	01

Таблица 6

Кодирование внутреннего состояния автомата U

Элементарные состояния	U_2	U_1
0	0	0
1	0	1
2	1	1
3	1	0

Проведя структурный синтез автоматов на триггерных элементах, запишем функции возбуждения для триггеров с отдельными входами (q_{01} , u_{01} , q_{10} , u_{10}) и для триггера со счетным входом (q_s , u_s). Обозначив через Q_i^0 (U_i^0) нулевое, а через Q_i^1 (U_i^1) единичное состояние i -го элементарного автомата, для автоматов Q_1 , U_1 будем иметь:

$$q_{01}^1 = x_2 \vee x_4',$$

$$u_{01}^1 = x_2,$$

$$q_{10}^1 = Q_2^0 x_6,$$

$$u_{10}^1 = x_4,$$

$$q_s^1 = x_2 \vee x_4' \vee Q_2^0 x_6,$$

$$u_s^1 = x_2 \vee U_1^1 x_4.$$

Соответственно для элементарных автоматов Q_2 , U_2 имеем:

$$q_{01}^2 = x_4' \vee x_5,$$

$$u_{01}^2 = x_3,$$

$$q_{10}^2 = x_6,$$

$$u_{10}^2 = x_4,$$

$$q_s^2 = x_4' \vee x_5 \vee Q_2^1 x_6,$$

$$u_s^2 = x_3 \vee U_2^1 x_4.$$

Нетрудно видеть, что функции возбуждения для Q_2 , U_1 , U_2 получаются простыми, если в качестве элементарного автомата выбрать триггер с отдельными входами, а для Q_1 это безразлично. В качестве Q_1 мы выбрали также триггер с отдельными входами.

На основании таблицы выходов и таблицы переходов можно записать функции выходных сигналов:

$$y_1 = x_4' \vee U_2^0 x_5,$$

$$y_2 = Q_2^0 Q_1^1 x_3 \vee Q_2^1 U_2^1 U_1^1 x_6,$$

$$y_3 = Q_2^0 U_1^0 x_3 \vee U_2^1 U_1^0 x_6,$$

$$y_4 = Q_2^0 x_4' \vee U_2^1 x_5 \vee Q_2^1 Q_1^1 U_1^0 x_6,$$

$$y_5 = Q_2^0 U_1^0 x_6,$$

где

$$x'_4 = U_1^1 x_4 \text{ и } x'_4 = U_1^0 x_4.$$

Рассмотрим теперь временные характеристики для синхронизации работы УАПС с УЦУ по импульсно-временной диаграмме, изображенной на рис. 6. Если обозначить через t интервал времени между поступающими сигналами x_5 и x_6 (т. е. время одного цикла УЦУ при считывании или записи), то для правильного функционирования системы необходимо, чтобы было

$$t < n\tau - \delta,$$

где τ — интервал между маркерными синхронизирующими импульсами,

$n = 30$ — разрядность регистра при работе со словарем,

δ — величина, учитывающая нестабильность скорости вращения барабана.

Если принять, что

$$\delta \approx (0,1 + 0,2) n\tau,$$

то получим:

$$t \leq 0,8 n\tau.$$

Запись информационной части статьи в ДОЗУ производится по $n = 30$ разрядов, для чего необходимо, чтобы время одного цикла работы УЦУ при записи также не превышало t . Если выбрать $\tau = 5$ мксек, то для t получим условие $t < 120$ мксек.

В заключение приведем некоторые технические характеристики барабанов, использованных в машине «Гарни».

1. Тип МБ — В1 с плавающими в воздухе головками.
2. Количество информации на одном МБ = 500 000 двоичных знаков, при плотности записи 10 имп/мм.
3. Скорость вращения 1450 об/мин (время одного оборота ~ 40 мсек).
4. Число двоичных знаков по окружности барабана — 8892.
5. Максимальная частота следования импульсов — 250 кгц.

Ն. Վ. ՎԱՐԴԱՆՅԱՆ, Է. Ս. ԵՂԻԱՋԱՐՅԱՆ, Ռ. Լ. ՌԻՌԻՏՅԱՆ

«ԳԱՌՆԻ» ՄԵՔԵՆԱՅԻ ԲԱՌԱՐԱՆՆԵՐԻ ԿԱԶՄԱԿԵՐՊՈՒՄԸ

Ա մ փ ո փ ու մ

Բառարանների կազմակերպումը և փնտրումը բառարաններում յուրահատուկ տեղ են գրավում մեքենայական թարգմանության մեջ, որովհետև համընդհանուր հաշվիչ մեքենաների օգնությամբ կատարվող մեքենայական թարգմանության փորձերը ցույց են տալիս, որ նախադասության թարգմանության

համար անհրաժեշտ ժամանակի մեծ մասը ծախսվում է նրա մեջ մտնող բառերի թարգմանության վրա: Այդ իսկ տեսակետից «Գառնի» մեքենայում հատուկ ուշադրություն է դարձված բառարանների կազմակերպմանը:

Հոդվածում նկարագրված են մեքենայի բառարանների կազմակերպման և ասոցիատիվ փնտրման սկզբունքները, որոնց շնորհիվ՝

1) Նախադասության թարգմանության համար անհրաժեշտ ժամանակի շատ բիշ մասն է ծախսվում բառերի թարգմանության վրա:

2) Բառարանների գրանցման համար հատկացվում են համեմատաբար ավելի բիշ հիշող էլեմենտներ (նիշեր):

Ուսումնասիրվում են բառերի փնտրման համար անհրաժեշտ ղեկավարական սխեմաների կառուցման հետ կապված հարցերը:

Л И Т Е Р А Т У Р А

1. И. Л. Братчиков, С. Я. Фитиалов, Г. С. Цейтин. О структуре словаря и кодировке информации для машинного перевода. Материалы по МП, сб. 1, ЛГУ, 1958.
2. Л. А. Калужнин, А. А. Стогний, Л. С. Стойкова. О принципах построения машинных словарей. «Прикладная лингвистика и машинный перевод», Изд. Киевского ун-та, 1962.
3. З. Л. Рабинович, А. А. Стогний, И. Г. Мороз. Об одном варианте построения информационного устройства на магнитном барабане. «Вопросы вычислительной техники» (машины, устройства, элементы и их применение), Киев, Гос. изд. тех. лит. УССР, 1961.
4. О. С. Кулагина, Г. В. Вакуловская. Опытные переводы с французского языка на русский на машине «Стрела». «Проблемы кибернетики», вып. 2, М., 1959.
5. Н. В. Вартанян и Э. В. Егиазарян. Организация словарей с ассоциативным поиском. «Тезисы докладов конференции по машинному переводу» (Ереван, 17—22 апреля 1967 г.).
6. Р. А. Базмаджян, Т. К. Джанполадян, Т. М. Тер-Микаэлян. Об одном варианте морфологического анализа и методе его реализации на специализированной машине. «Проблемы кибернетики», вып. 15, М., 1965.
7. Е. Н. Вавилов и Г. П. Поргной. Синтез схем электронных цифровых машин. «Советское радио», 1963.
8. Е. Н. Вавилов, Д. Б. Шишков. Об одном методе экономического кодирования автоматов. «Техническая кибернетика», 1967, № 2.
9. В. Zacharov. A refinement in coding the Russian Cyrillic alphabet—MT, 1957, 4, № 3.