

УДК 681.3.068

Е. Г. КАЧКО, канд. техн. наук, *Ю. С. МАРЧЕНКО*

**СИСТЕМА ПРОГРАММИРОВАНИЯ ДАННЫХ ИССЛЕДОВАНИЯ
ВКУСОВОГО АНАЛИЗАТОРА
СООБЩЕНИЕ 2**

В работе [1] предложен специальный язык для описания объективных данных обследования вкусового анализатора человека и показана его применимость для задания данных, полученных общим и капельным методами и методом плетизмографии. Сообщение 2 посвящено описанию транслятора с этого языка.

При разработке транслятора используются синтаксические методы трансляции [2]. Транслятор может быть представлен в виде суперпозиции трех блоков: блок лексического анализа (БЛА); блок синтаксического анализа (БСА); блок формирования шкалы (БФШ).

Будем называть текст программы, написанной на входном языке, исходным текстом. Понятие лексической единицы (ЛЕ) совпадает с общепринятым [2]. БЛА предназначен для анализа исходного текста с целью построения таблицы чисел, S-строки и обнаружения лексических ошибок, допущенных в записи ЛЕ. БСА предназначен для проверки корректности исходного текста с точки зрения синтаксиса. При нарушении необходимой последовательности ЛЕ формируются соответствующие диагностические сообщения. БФШ ставит в соответствие данные обследования из исходного текста данным таблицы [1] для

определения уровней отклонений от нормы в каждом конкретном случае.

Входной информацией для БЛА являются исходный текст; таблица терминальных символов.

Терминальные символы вместе с соответствующими кодами, записанными в 16-ричной системе счисления, представлены в табл. 1. В эту же таблицу занесены ЛЕ типа «понятие», «число» с соответствующими кодами.

Таблица 1

Терминальный символ	16-ричный код	№ п/п	Терминальный символ	16-ричный код
Общий метод	20	16	Винная кислота	2F
Левая сторона, капельный метод	21	17	Соляная кислота	30
Правая сторона, капельный метод	22	18	Хинин солянокислый	31
Плетизмографии метод	23	19	Понятие (имя болезни)	40
Сладкое	24	20	Пробел ()	01
Кислое	25	21	Запятая (,)	02
Соленое	26	22	Точка с запятой (;)	03
Горькое	27	23	Символ разделения	04
Появление реакции	28	24	Звездочка (*)	05
Сдвиг максимальный	29	25	Двоеточие (:)	06
Длительность реакции	2A	26	*L	07
Величина реакции	2B	27	*E	08
Сахар	2C	28	Число	09
Глюкоза	2D	29	Имперсанд (&)	0A
Хлористый натрий	2E			

Выходной информацией блока являются: таблица чисел (TABN); S-строка (STR); список диагностических сообщений, если в исходном тексте допущены лексические ошибки (ERR).

Таблица чисел представляет собой последовательность чисел в машинной форме в том порядке, в котором они встречаются в исходном тексте. Если одно и то же число появляется несколько раз, в таблицу чисел оно записывается только один раз. Каждое число рассматривается как константа с плавающей точкой с обычной точностью вне зависимости от его внешнего представления.

S-строка — это таблица стандартных символов, под каждый элемент которой отводится слово (4 байта). Структура элемента S-строки представлена на схеме. Самый старший байт всегда заполняется шестнадцатиричными CØ, что является признаком S-строки. Код ЛЕ выбирается из табл. 1. Для чисел в качестве дополнительной информации заносится номер числа в таблице чисел, для ЛЕ другого типа — константа, содержащая нули во всех разрядах. Каждой ЛЕ исходного текста ставится в соответствие один элемент S-строки. Последователь-

ность ЛЕ и соответствующих им элементов в S-строке одинакова.

С_Q | Код ЛЕ | Дополнительная информация
 Байт 0 | Байт 1 | Байты 2,3

Т а б л и ц а 2

$i \backslash j$	1	2	3	4	5	6	7	8	9	10	11	12	Примечание
1		=											Понятие
2			=										*L
3							=						Метод
4							=						Компонента
5							=						Эталон
6								=	=				Число
7				=	=	=							:
8				=									,
9			=							=			:
10	=										=		*
11													&
12													Временные характеристики

Диагностические сообщения формируются в виде СТРОКА (№ строки) (текст сообщения).

Одно диагностическое сообщение занимает 120 позиций, что соответствует одной строке АЦПУ, № строки совпадает с номером строки исходного текста. Программа просматривается с начала до конца, диагностические сообщения заносятся в область ошибок (ERR). Распечатка области ошибок ERR выполняется после отработки БЛА и БСА с целью выявления максимального количества ошибок за один просмотр.

Входной информацией для БСА являются: S-строка (STR); матрицы предшествования (M).

Матрицы предшествования представляют собой таблицы [3], в которых указаны допустимые и недопустимые последовательности стандартных символов S-строки. Ввиду простоты операторов мы использовали упрощенный вид матрицы предшествования. В табл. 2 представлена матрица предшествования для

общего и капельного методов, а в табл.3 — для метода плетизмографии. Элемент матрицы M_{ij} обозначается $\overset{\cdot}{=}$, если стандартный символ j может следовать за стандартным символом i . Отсутствие этого знака означает недопустимость следования j -символа после i -го.

Т а б л и ц а 3

$i \backslash j$	1	2	3	4	5	6	7	8	9	10	11	12	Примечание
1		$\overset{\cdot}{=}$											Понятие
2			$\overset{\cdot}{=}$										*L
3							$\overset{\cdot}{=}$						Метод
4							$\overset{\cdot}{=}$						Компонента
5							$\overset{\cdot}{=}$						Эталон
6								$\overset{\cdot}{=}$	$\overset{\cdot}{=}$				Число
7				$\overset{\cdot}{=}$	$\overset{\cdot}{=}$	$\overset{\cdot}{=}$						$\overset{\cdot}{=}$:
8				$\overset{\cdot}{=}$								$\overset{\cdot}{=}$,
9			$\overset{\cdot}{=}$							$\overset{\cdot}{=}$;
10	$\overset{\cdot}{=}$										$\overset{\cdot}{=}$		*
11													&
12							$\overset{\cdot}{=}$						Временные характеристики

Выходной информацией БСА являются: список диагностических сообщений, если в исходном тексте допущены синтаксические ошибки (нарушены правила следования ЛЕ); признак успешного завершения БЛА и БСА.

Диагностические сообщения о синтаксических ошибках формируются в виде
ЗАПИСЬ (№ записи) НЕ УКАЗАН МЕТОД ОБСЛЕДОВАНИЯ
 или

ЗАПИСЬ (№ записи) СИНТАКСИЧЕСКАЯ ОШИБКА

Диагностическое сообщение первого типа указывает на отсутствие метода обследования, а второго — на нарушение правил следования ЛЕ. Запись — это последовательность символов исходного текста до ближайшего символа «;». Все диагностические сообщения заносятся в область *ERR* вслед за сообщениями БЛА.

Признак успешного завершения

$$PR = \begin{cases} 1, & \text{отсутствие ошибок в исходном тексте;} \\ 0, & \text{наличие ошибок.} \end{cases}$$

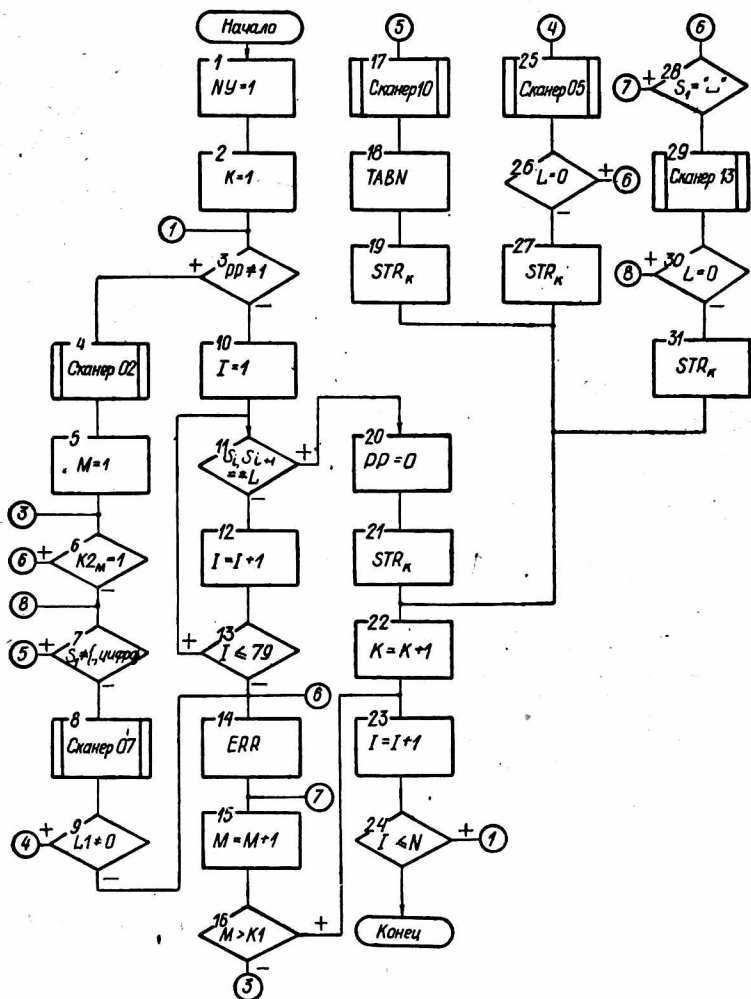


Рис. 1

Если $PR = 0$, то после распечатки области ERR работа транслятора прекращается. БФС выполняется только при $PR = 1$. Входной информацией БФС являются таблица чисел ($TABN$): S -строка (STR); таблица исходных данных (TAB).

Таблица исходных данных содержит шкалу для определения отклонений данных обследования вкусового анализатора от нормы [1] для различных методов, компонент, эталонов и параметров времени.

Выходной информацией БФС является массив *A*, элементы которого указывают степень отклонения от нормы отдельных компонент при выбранных методе исследования и эталонах. Степень отклонения от нормы, оцененная по пятибалльной системе, является существенным признаком для определения характера заболевания, его стадии и прогнозов.

При выполнении БФС может быть сформировано диагностическое сообщение.

ЗАПИСЬ (№ записи) ОШИБКА В ЗАДАНИИ ЭТАЛОНА.

При отсутствии ошибок после выполнения блока лексического анализа, блока синтаксического анализа и блока формирования шкалы печатается сообщение

ТРАНСЛЯЦИЯ УСПЕШНАЯ.

Опишем функционирование отдельных блоков транслятора.

После ввода исходного текста он распечатывается в виде

НОМЕР	НОМЕР	
СТРОКИ	ЗАПИСИ	(текст исходной программы)
1	1	
...

Исходный текст после распечатки приводится к эталонному виду — все ключевые слова сокращаются до четырех символов включительно. Затем выполняется БЛА. Блок-схема БЛА приведена на рис. 1. Обозначения, принятые на блок-схеме:

NY — номер записи;

PP — признак понятия

$$PP = \begin{cases} 1, & \text{работа в режиме обучения} \\ 0, & \text{работа в режиме постановки диагноза;} \end{cases}$$

I — номер строки исходного текста;

K — номер элемента в *S*-строке, соответствующего ЛЕ;

M — номер ЛЕ;

K1 — количество ЛЕ;

K2 — количество символов в ЛЕ;

S — массив символов ЛЕ;

i — номер символа в массиве *S*;

L — признак наличия или отсутствия ошибки в записи ЛЕ;

$$L = \begin{cases} 0, & \text{ошибка есть,} \\ 1, & \text{ошибки нет;} \end{cases}$$

STR, *TABN* — *S*-строка и таблица чисел соответственно;

ERR — область ошибок;

СКАНЕРØ2 — подпрограмма разбивки исходного текста на ЛЕ;

СКАНЕРØ7 — подпрограмма анализа правильности записи числа;

СКАНЕРØ1Ø — подпрограмма перевода исходного числа в форме представления с плавающей точкой;

СКАНЕРØ5 — подпрограмма анализа правильности записи ключевых слов;

N — количество строк исходного текста.

Описание блок-схемы.

Блоки 1, 2 — установка начальных значений *NY*, *I*, *K*;

- Блок 3 — проверка режима работы системы;
 Блоки 10—13 — проверка первой строки на название болезни, если режим обучения;
 Блок 14 — формирование диагностического сообщения при отсутствии названия болезни или программами анализа чисел и служебных слов;
 Блок 4 — разделение строки исходного текста на ЛЕ;
 Блоки 5, 15, 16 — перебор лексических единиц строки и анализ на полноту перебора;
 Блоки 7—9, 17—19 — проверка очередной ЛЕ на число и формирование элемента S-строки и TAVN (при необходимости);
 Блоки 20—21 — формирование признака PP и элемента S-строки в случае режима обучения;
 Блоки 25—27 — проверка очередной ЛЕ на ключевое слово и формирование элемента S-строки, если необходимо;
 Блоки 28—31 — проверка очередной ЛЕ на пробел и разделитель и формирование очередного элемента S-строки, если необходимо;
 Блоки 23—24 — переход к очередной строке исходного текста и проверка на его окончание.

При работе БЛА возможно формирование следующих диагностических сообщений:

СТРОКА (№ строки) ЛЕ НЕ ОПРЕДЕЛЕНА
 СТРОКА (№ строки) ОТСУТСТВУЕТ НАЗВАНИЕ БОЛЕЗНИ
 СТРОКА (№ строки) В ЧИСЛЕ ЗАПРЕЩЕННЫЙ СИМВОЛ
 СТРОКА (№ строки) В ЧИСЛЕ БОЛЕЕ ОДНОЙ ТОЧКИ
 СТРОКА (№ строки) В МАНТИССЕ НЕТ ЦИФР
 СТРОКА (№ строки) ОШИБКА В ЗАПИСИ ПОРЯДКА
 СТРОКА (№ строки) ПЕРЕПОЛНЕНИЕ ПОРЯДКА
 СТРОКА (№ строки) В МАНТИССЕ БОЛЬШЕ 7 ЦИФР

Блок-схема БСА приведена на рис. 2.

Обозначения, принятые на блок-схеме:

I — признак метода;

$$I = \begin{cases} 0 & \text{метод не задан.} \\ 1 & \text{общий или капельный метод,} \\ 2 & \text{метод плетизмографии;} \end{cases}$$

Π — массив, предназначенный для указания диапазона допустимых кодов для каждой строки матрицы предшествования;

M — матрица предшествования;

NY — номер записи;

I — номер стандартного символа в S-строке;

KODR — код текущего стандартного символа;

KOD — код очередного стандартного символа;

L1 (L2) номер строки (столбца) матрицы предшествования, соответствующих I (I + 1) стандартным символам S-строки;

L — текущее значение номера строки (столбца) матрицы предшествования;

ERR — область ошибок;

NERR — величина, определяющая адрес очередного диагностического сообщения;

СКАН 21 — подпрограмма для формирования матриц предшествования M (2, 12, 12);

СКАН 23 — подпрограмма для выделения кода стандартного символа S-строки;

СКАН 24 — подпрограмма для формирования допустимого диапазона изменения кодов $\Pi(12,2)$ для каждой строки матриц предшествования;

ERR1 — диагностическое сообщение

ЗАПИСЬ (№ записи) НЕ УКАЗАН МЕТОД ОБСЛЕДОВАНИЯ;

ERR2 — диагностическое сообщение

ЗАПИСЬ (№ записи) СИНТАКСИЧЕСКАЯ ОШИБКА.

Описание блок-схемы. После формирования диапазона допустимых кодов (1) для каждой строки матриц предшествования, самих матриц и задания начального значения номера записи NY (блоки 1—4) начинается просмотр стандартных символов S -строки (блоки 5—6). Если код стандартного символа равен 64, т. е. соответствует названию болезни, то переходим к стандартному

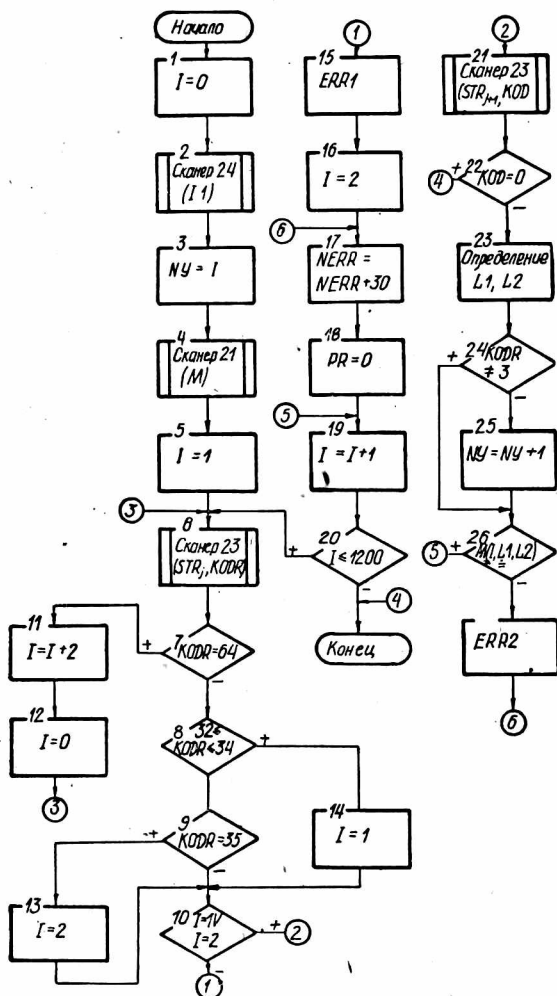


Рис. 2

символу, соответствующему методу обследования (блоки 7—12). Если код не равен 64, то проверяем, соответствует ли очередной символ методу обследования, если «нет», то $ERR1$, если «да», то определяем номер метода ($I = 1$ или $I = 2$) (блоки 7—10, 13—16). При наличии $ERR1$ предполагаем, что $I = 2$, признак наличия ошибок PR устанавливаем в 0 и переходим к анализу очередного стандартного символа (блоки 16—20). При отсутствии $ERR1$ после

проверки на конец S-строки (блок 22) определяем номер строки и столбца в матрице предшествования M , соответствующих j и $j+1$ стандартным символам (блок 23). Одновременно, если STR соответствует символу «;», то изменяем номер записи NY (блок 25).

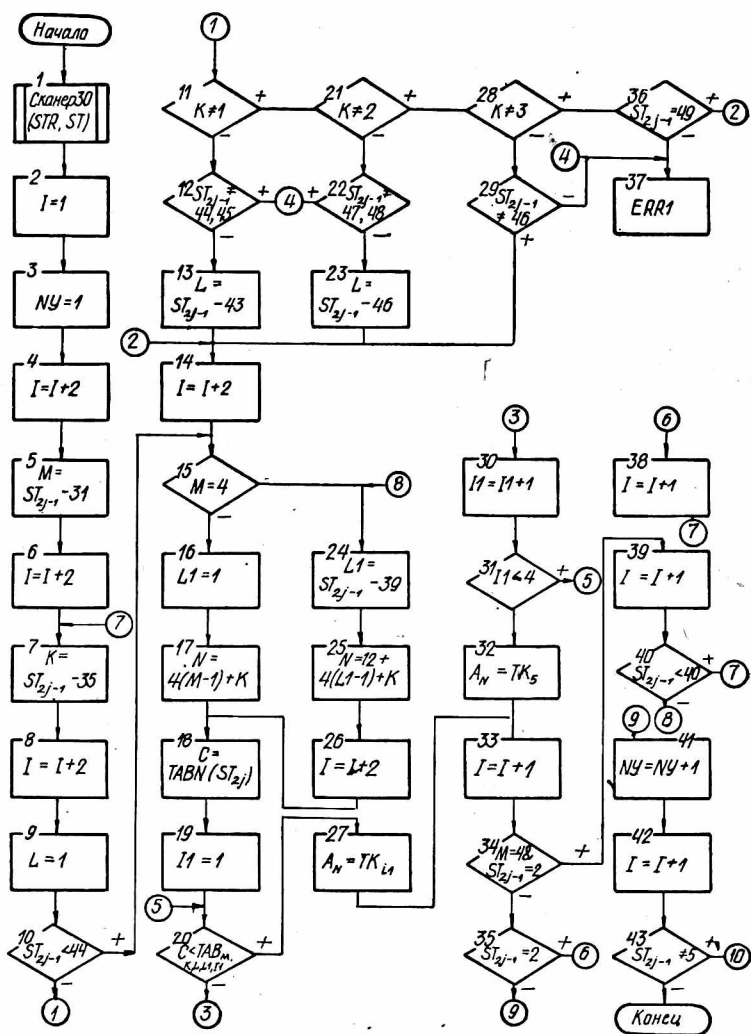


Рис. 3

Если соответствующий элемент матрицы предшествования равен «истине», что соответствует символу $\underline{=}$, то переходим к анализу очередной пары символов (блоки 26, 19—20). если «лжи», что соответствует отсутствию символа $\underline{=}$, то формируется $ERR2$ (блок 6) и выполняется переход к блокам для обработки ошибок (блоки 17—18).

Просмотр элементов S -строки завершается, если достигнут признак конца S -строки ($KOD = 0$); обработаны все элементы S -строки ($j_{max} = 1200$).

Блок-схема БФС приведена на рис. 3.

Обозначения, принятые на блок-схеме:

- j — номер элемента S -строки;
 STR — S -строка;
 $KOD(1)$ — код стандартного символа S -строки;
 $KOD(2)$ — номер числа в таблице чисел $TABN$;
 ST — имя «распакованной» S -строки;
 M — номер метода

$$M = \begin{cases} 1, & \text{общий метод,} \\ 2, & \text{левая сторона, капельный метод,} \\ 3, & \text{правая сторона, капельный метод,} \\ 4, & \text{метод плетизмографии;} \end{cases}$$

K — номер компоненты

$$K = \begin{cases} 1, & \text{сладкое,} \\ 2, & \text{кислое,} \\ 3, & \text{соленое,} \\ 4, & \text{горькое;} \end{cases}$$

L — номер эталона

$$L = \begin{cases} 1 & \text{сахар, винная кислота, соляная кислота,} \\ & \text{хинин солянокислый,} \\ 2 & \text{глюкоза, соляная кислота;} \end{cases}$$

NY — номер записи;

A — выходной массив, состоящий из элементов типа $LTLT$, $LESS$, ...

N — номер элемента в массиве A ($N = 1, 2, \dots, 28$);

TAB — таблица исходных данных, размерность массива равна 5, первый индекс определяет номер метода (M),

второй — номер компоненты (K),

третий — номер эталона (L),

четвертый — параметр времени ($L1$)

пятый — номер интервала ($I1$);

$I1$ — номер интервала в TAB ($I1 = 1, 2, \dots, 5$);

C — число, значение которого определяется из таблицы чисел;

TK — массив кодов (уровней отклонений от нормы $LTLT$, $LESS$, $NORM$, $MORE$, $GTGT$);

$L1$ — номер компоненты времени для метода плетизмографии

$$L1 = \begin{cases} 1, & \text{появление реакции,} \\ 2, & \text{едвиг максимальный,} \\ 3, & \text{длительность реакции,} \\ 4, & \text{величина реакции;} \end{cases}$$

$ERR1$ — диагностическое сообщение

ЗАПИСЬ (№ записи) ОШИБКА В ЗАДАНИИ ЭТАЛОНА.

Описание блок-схемы.

Блок 1 — «распаковка» S -строки, т. е. запись кода и номера числа в $TABN$ для каждого символа S -строки в слово;

Блоки 2 и 3 — задание начальных значений I , NY ;

Блоки 4 и 5 — определение метода;

Блоки 6—13, 21—23, 28, 29, 36 — определение номера компоненты;

Блок 37 — формирование диагностического сообщения $ERR1$, и аварийного выхода из подпрограммы (КОНЕЦ!);

Блоки 16 и 17 — формирование индекса элемента в массиве A для общего и капельного методов;

Блоки 24—26 — формирование индекса элемента в массиве A для метода плетизмографии;

Блоки 18—20, 27, 30—32 — формирование элемента массива A;
 Блоки 33—35, 38—40 — переход к анализу новой компоненты, если она есть
 Блоки 41—43 — переход к анализу нового метода, если он есть
 и успешный выход из подпрограммы, если все методы исчерпаны.

При разработке транслятора программы собирались и тестировались снизу вверх. Ниже приведены примеры тестирования транслятора для случаев отсутствия и наличия ошибок в исходном тексте.

Пример 1.

НОМЕР СТРОКИ	НОМЕР ЗАПИСИ	
1	1	САХАРНЫЙ ДИАБЕТ *L
2	1	ОБЩИЙ МЕТОД:
3	1	КИСЛОЕ: ВИННАЯ КИСЛОТА: 0.01,
4	1	ГОРЬКОЕ: 0.00007,
5	1	СЛАДКОЕ: ГЛЮКОЗА: 1,5,
6	1	СОЛЕНОЕ: ХЛОРИСТЫЙ НАТРИЙ: 0.7;
7	2	*

GTGT NORM MORE NORM
 ТРАНСЛЯЦИЯ УСПЕШНАЯ

Этот пример соответствует случаю отсутствия ошибок в исходном тексте.

ПРИМЕР 2.

НОМЕР СТРОКИ	НОМЕР ЗАПИСИ	
... 1	1	ЗАПИСЬ С ОШИБКАМИ *L
2	1	ОБЩИЙ МЕТОД:
3	1	ГОРЬКОЕ: ГЛЮКОЗА: 0,358,
4	1	КИСЛОЕ: 3.4E—02,
5	1	СЛАДКОЕ: ХЛОРИСТЫЙ НАТРИЙ: 0.024,
6	1	СОЛЕНОЕ: 0.038E + 1;*

ДИАГНОСТИЧЕСКИЕ СООБЩЕНИЯ
 ЗАПИСЬ 1 ОШИБКА В ЗАДАНИИ ЭТАЛОНА

Система обнаружила ошибку, связанную с неправильным заданием эталонов для ГОРЬКОЕ и СЛАДКОЕ.

ПРИМЕР 3.

НОМЕР СТРОКИ	НОМЕР ЗАПИСИ	
1	1	ПРОВЕРКА ДИАГНОСТИЧЕСКИХ СООБЩЕНИЙ *L
2	1	МЕТОД ПРОБ: 0.00 000 065, 65 000 000;
3	2	СЛАДКОЕ: 0.023E 1,
4	2	СЛАДКОЕ: 5.37,
5	2	КИСЛОЕ: 5.A31;
6	3	ГОРЬКОЕ: A.531;
7	4	СОЛИНОЕ: 0,537.01;

ДИАГНОСТИЧЕСКИЕ СООБЩЕНИЯ
 СТРОКА 2 ЛЕ НЕ ОПРЕДЕЛЕНА
 (МЕТОД ПРОБ не может быть распознан)
 СТРОКА 2 В МАНТИССЕ БОЛЬШЕ 7 ЦИФР
 СТРОКА 2 В МАНТИССЕ БОЛЬШЕ 7 ЦИФР
 (оба числа, записанные в строке 2, содержат больше 7 цифр в мантиссе)
 СТРОКА 5 В ЧИСЛЕ ЗАПРЕЩЕННЫЙ СИМВОЛ
 (5.A31 содержит запрещенный символ A)
 СТРОКА 7 Л. Е. НЕ ОПРЕДЕЛЕНА
 (ключевое слово СОЛИНОЕ написано с ошибкой)
 СТРОКА 7 В ЧИСЛЕ БОЛЕЕ ОДНОЙ ТОЧКИ
 ЗАПИСЬ 1 НЕ УКАЗАН МЕТОД ОБСЛЕДОВАНИЯ

ЗАПИСЬ 1 СИНТАКСИЧЕСКАЯ ОШИБКА
ЗАПИСЬ 2 СИНТАКСИЧЕСКАЯ ОШИБКА
ЗАПИСЬ 2 СИНТАКСИЧЕСКАЯ ОШИБКА
ЗАПИСЬ 2 СИНТАКСИЧЕСКАЯ ОШИБКА
ЗАПИСЬ 3 СИНТАКСИЧЕСКАЯ ОШИБКА
ЗАПИСЬ 3 СИНТАКСИЧЕСКАЯ ОШИБКА
ЗАПИСЬ 4 СИНТАКСИЧЕСКАЯ ОШИБКА
ЗАПИСЬ 4 СИНТАКСИЧЕСКАЯ ОШИБКА

Некоторые из указанных синтаксических ошибок являются следствием ошибок в задании ЛЕ, т. е. прежде всего необходимо исправить лексические ошибки.

Транслятор занимает 10638 байтов, для реализации БЛА использован язык АССЕМБЛЕР, а БСА и БФС — язык ФОРТРАН.

Таким образом, данные обследования вкусового анализатора, заданные на языке, близком к общепринятому, переводятся на машинный язык. Эти данные используются для обучения диагностических систем и уточнения диагноза в процессе их функционирования. В ходе трансляции отыскиваются ошибки в задании данных. Специальные диагностические сообщения позволяют локализовать эти ошибки и определить их характер.

Список литературы: 1. Качко Е. Г., Марченко Ю. С. Система программирования данных исследования вкусового анализатора. *Сообщение 1.* — Проблемы бионики, 1982, вып. 28, с. 118—121. 2. Donovan Дж. Системное программирование. — М.: Мир, 1975, с. 296—352. 3. Лебедев В. Н. Введение в системы программирования. — М.: Статистика, 1975, с. 232—253.

Поступила в редколлегию 02.03.81.