

Ц 8405

Г-15

СООБЩЕНИЯ
ОБЪЕДИНЕННОГО
ИНСТИТУТА
ЯДЕРНЫХ
ИССЛЕДОВАНИЙ
ДУБНА



24/11-78

1879/2-78

Ю - 11228

В.В.Галактионов

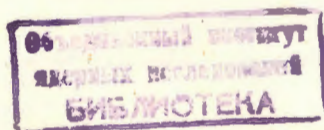
**СИНТАКСИЧЕСКИЙ АНАЛИЗ
И ТРАНСЛЯЦИЯ КОМАНД
КОНЦЕНТРАТОРА ТЕРМИНАЛОВ**

1978

10 - 11228

В.В.Галактионов

СИНТАКСИЧЕСКИЙ АНАЛИЗ
И ТРАНСЛЯЦИЯ КОМАНД
КОНЦЕНТРАТОРА ТЕРМИНАЛОВ



Галактионов В.В.

10 - 11228

Синтаксический анализ и трансляция команд концентратора терминалов

Создание транслятора с языка команд, способ задания синтаксиса и метод трансляции которых предлагаются в данной работе, является частью общей разработки математического обеспечения концентратора терминалов при БЭСМ-6, создаваемого на базе ЭВМ ЕС-1010.

В качестве входного языка команд концентратора принят язык системы INTERCOM для ЭВМ серии 6000 фирмы CDC (в английском и русском вариантах).

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1978

Galaktionov V.V.

10 - 11228

Syntax Analysis and Translation of Terminal Concentrator Commands

Methods for creating a translator from command languages, for defining the syntax and translation are suggested. The work presented is a part of general development of software for terminal concentrator for the BESM-6 computer which is created on the base of the ES-1010 computer. The language of INTERCOM system for computers 6000 series of CDC firm (in English and Russian versions) serves as an input language for command concentrator.

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

Communication of the Joint Institute for Nuclear Research. Dubna 1978

ВВЕДЕНИЕ

Данная работа является частью общей разработки системы математического обеспечения многопультного концентратора, создаваемого на основе ЭВМ ЕС-1010.

В качестве языка команд концентратора принят язык системы INTERCOM для ЭВМ серии 6000 фирмы CDC (в английском и русском вариантах). Любая набираемая пользователем на клавиатуре дисплея и поступающая в ЭВМ текстовая строка рассматривается как команда, либо как очередная строка формируемого текстового файла. При анализе принятой команды возникает проблема грамматического разбора входного текста, трансляции и выполнения команды. Процесс разбора или анализа строки исходного текста заключается в нахождении соответствия этого типа определенному синтаксическому правилу принятого языка. Эту задачу должен выполнять анализатор, реализующий алгоритм синтаксического анализа.

Эффективность синтаксического разбора в высшей степени зависит от способа задания грамматики входного языка. Например, по заданным произвольным способом контекстно-свободным грамматикам, каковыми являются грамматики большинства искусственных языков (языков программирования и диалоговых), нельзя однозначно определить алгоритм синтаксического разбора, а можно лишь выделить его возможные пути (ветви). При этом синтаксический анализ обычно состоит в проверке: не приводит ли выбранный вариант разбора к тупиковой ситуации. Если это случается, происходит возврат к точке разветвления и проверяется следующая ветвь разбора и т.д. Таким образом, наличие множества тупиковых ветвей разбора и трудностью выбора оптимального пути анализа объясняется во многих случаях невысокая эффективность работы синтаксически-ориентированных трансляторов, что, в конечном итоге, как правило,

является следствием неудачного способа задания грамматики входного языка. Поэтому основными приемами повышения эффективности и ускорения синтаксического разбора входных текстов являются приведение исходной грамматики (эквивалентное преобразование) к грамматике, обладающей новыми свойствами (например, превращение исходной контекстно-свободной грамматики в контекстно-чувствительную путем добавления в каждое исходное синтаксическое правило необходимого левого и правого контекста), либо поиски новых форм ее задания.

В данной работе предлагается способ формального задания синтаксиса (метаязык) входного языка, имеющий высокую скорость распознавания входных текстов и определяющий однозначный путь грамматического разбора. Кроме того, приводится описание синтаксического анализатора, осуществляющего вместе с грамматическим разбором входного текста его трансляцию. Система обработки входных команд является собой синтаксически-ориентированный транслятор. Анализирующая часть его зависит не от входного языка, а лишь от способа его задания (метаязыка). Переход к новому языку может производиться путем замены синтаксиса языка (синтаксических таблиц), описываемого средствами предлагаемого метаязыка.

Метаязык

В отличие, например, от метаязыка Бэкуса в предлагаемом метаязыке для задания синтаксиса входного языка используются правила следования, а не правила подстановки; кроме того, в метаязыке вводится понятие памяти магазинного типа. Принцип магазина в данном случае используется лишь для автоматической продвижки элементов в одном направлении (выборка). Вводятся следующие операции над памятью:

- установить магазин,
- выбрать элемент,
- выполнить команду. В данном случае выбирается элемент из магазина и интерпретируется как команда одного из типов:

1. Сравнение (CP). Команда содержит код символа и адрес. В случае сравнения кода этого символа с кодом заданного символа происходит переход по адресу. В случае несовпадения - выборка следующего элемента магазина и выполнение новой команды и т.д.

2. Переход по магазину (PM). Выбранный элемент представляет собой безусловную команду перехода.

В обоих случаях (переход по сравнению и безусловный переход) выполнение команды перехода означает установку нового магазина с начальным адресом, равным адресу в выполняемой команде перехода.

Кроме того, в грамматику входного языка вводится еще один тип терминальных элементов - действия или семантики. Введение элементов этого типа предназначается для проведения трансляции языковых конструкций в процессе грамматического разбора. Предложение, порожденное грамматикой с действиями, будет содержать смесь слов (совокупности терминальных символов) и действий, однако грамматическому разбору будет подлежать предложение, состоящее только из слов. В описываемом метаязыке действия определяют номер семантической подпрограммы и являются указателями синтаксическому анализатору для формирования обращений к подпрограммам с этими номерами.

Таким образом, грамматикой предлагаемого входного языка будет следующее множество:

$$G = \{V, T, D, \Phi, A\}, \text{ где}$$

- V - словарь всех допустимых символов,
- T - множество терминальных символов,
- D - множество терминальных символов типа "действие",
- Φ - множество синтаксических правил (правил следования),
- A - аксиома грамматики.

Выполняется следующее соотношение $D \subset T \subset V$. Здесь не приводится детального описания множеств V, T и D, поскольку входной язык системы концентратора в настоящее время продолжает развиваться и, по-видимому, его алфавит будет значительно расширен.

Задание синтаксических правил

Как выше уже отмечалось, правилами описания синтаксиса данного языка являются правила следования. Это означает, что правые части каждого синтаксического правила представляют собой упорядоченные совокупности всех терминальных символов, относящихся к данному правилу. В этих правилах строго задается порядок следования (или варианты следований) элементов, что является важным обстоятельством для выбора однозначного пути грамматического разбора и его ускорения.

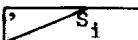
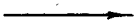
Для описания множества синтаксических правил Φ удобно использовать графический способ, при этом:

$\Phi = \{S, \Pi\}$, где

- $S = \{S_i\}$ - память магазинного типа (см. выше), в которой задается способ выбора синтаксического правила Π_j .
 Над каждым магазином S_i допускаются описанные выше операции (CP) и (ПМ).
 $\Pi = \{\Pi_j\}$ - множество правил следования. Правила задаются также в виде магазина, но к нему неприменимы никакие другие операции, кроме выборки элемента.

Магазин типа S может включать в себя элемент, называемый дном магазина, а правила следования Π - терминальные символы действия. Каждое синтаксическое правило оканчивается признаком конца всей языковой конструкции (предложения) либо разделителем. Разделитель содержит в себе код символа, выбранного в качестве знака разделителя, и указатель (адрес выбора) нового магазина S_k для продолжения грамматического разбора.

Введем обозначения:

- $a = a_1 a_2 \dots a_n$ - строка входного текста,
 $a(k)$ - элемент строки,
 \circ - указатель элемента магазина. Например,
 \textcircled{a} означает, что элементом является код буквы a.
 $>$, $<$ - угловые скобки определяют код "обобщенного" знака. Например, $\langle u \rangle$ означает, что элементом может быть любая десятичная цифра.
 Π_i^j - j-ый элемент i-го правила следования.
 S_i^j - j-ый элемент i-го магазина типа S.
 Π - задание терминального элемента "действие".
 i - номер семантической подпрограммы.
 $\#$ - "дно" магазина типа S.
 \downarrow - указатель конца предложения.
 - разделитель синтаксических правил.
 $"/>$ - знак разделителя, S_i - указатель нового магазина.
 - указатель следующего элемента в правиле следования.

Команда CP в магазине S задается в виде

○	Π _i
---	----------------

, где

○ содержит код символа для сравнения или признак безусловного (без сравнения) выбора правила, а Π_i - соответствующее синтаксическое правило.

Команда ПМ содержит условный код и адрес нового магазина S_k .

Код	S_k
-----	-------

В приложении 2 приведен пример графического задания с помощью принятых обозначений синтаксиса одной из команд редактора текстов для системы концентратора.

Синтаксический анализатор

Первый этап грамматического разбора и трансляции входной строки заключается в проведении лексического анализа текста - приведении к стандартному виду: редактировании и переводе во внутреннее представление. При этом устраняются "лишние" знаки, в частности, пробелы, заменяются всевозможные типы разделителей на один стандартный, вычисляется число значащих знаков и др. Редактирование входной строки не затрагивает специальных фрагментов - текстовых параметров, защищенных специальными знаками.

После лексического анализа входной строки собственно и начинается ее грамматический разбор, выполняемый программой синтаксического анализатора, входным параметром которой является адрес начала синтаксической таблицы, сформированной по описанным выше правилам. Это будет адрес первого магазина S_1 , содержащего информацию для выбора правила следования Π в зависимости от определяющих символов (одного или нескольких) входной строки.

В процессе анализа выполняются операции CP и ПМ над магазином типа S, выбирается очередное синтаксическое правило Π_i , заданное также в виде магазина, проверяется соответствие входной строки этому правилу и т.д.

При этом могут обнаруживаться синтаксические ошибки двух типов:

1) "Исчерпан" один из магазинов S_k (обнаружено "дно" магазина, что свидетельствует об обнаружении фрагмента исходного текста, к которому неприменимо ни одно из допустимых синтаксических правил.

2) Несоответствие языковой конструкции (или ее фрагмента) выбранному синтаксическому правилу.

Кроме того, при обнаружении в цепочке правил следования терминального элемента "действие", анализатор формирует обращение к соответствующей семантической подпрограмме, которая собственно и выполняет частичную (относящуюся к ней) трансляцию. При этом также могут обнаруживаться ошибки, но другого свойства - ошибки в семантике. При обнаружении ошибки грамматический разбор и трансляция входного текста прекращаются и в этом случае выходным параметром программы анализатора будет номер ошибки (номер текстовой строки для подпрограммы выдачи диагностики на терминал).

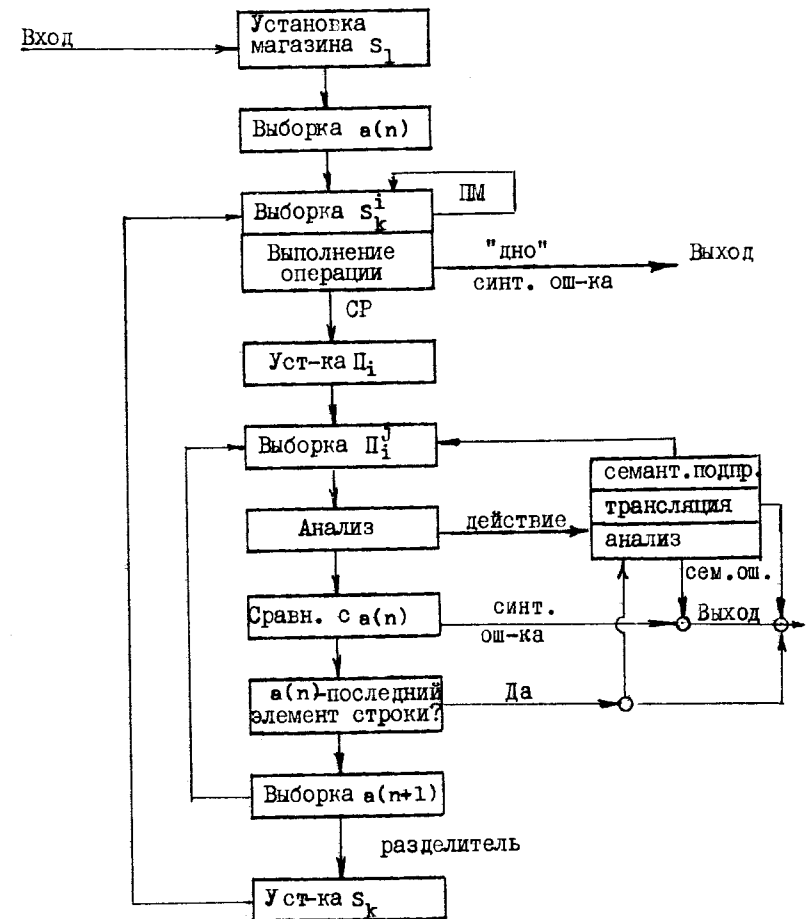
Таким образом, не считая лексического анализа, грамматический разбор и трансляция входной строки осуществляются за один просмотр текста. Результатом трансляции будет заполненная определенным образом таблица.

В приложении I приведена логическая схема работы синтаксического анализатора.

ЗАКЛЮЧЕНИЕ

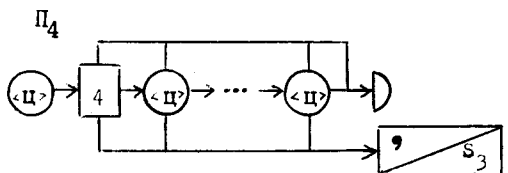
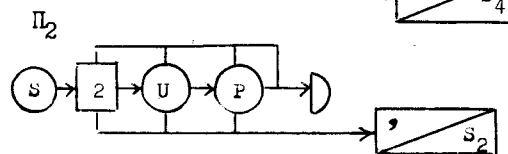
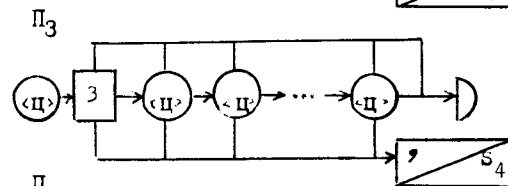
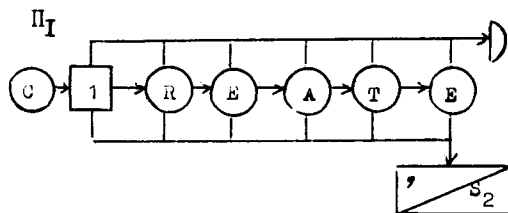
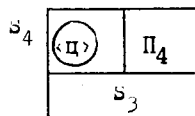
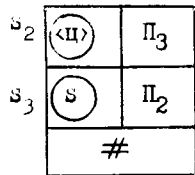
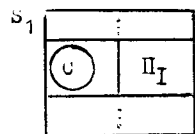
Описанный в данной работе алгоритм синтаксического анализа реализован автором на ЭВМ ЕС-1010 для грамматического разбора и трансляции входных команд языка диалогового общения пользователя за пультом терминала типа дисплея VT - 340 с вычислительной машиной в системе концентратора терминалов.

Транслятор работает в режиме интерпретации команд и запускается каждый раз по мере получения системой очередной команды пользователя.



Логическая схема работы машины синтаксического анализатора.

Приложение 2.



ЛИТЕРАТУРА

1. С.Я.Виленкин, Э.А.Трахтенгерц. Математическое обеспечение управляющих вычислительных машин. Изд-во "Энергия", Москва, 1972.
2. Дж.Фостер. Автоматический синтаксический анализ. Изд-во "Мир", Москва, 1975.
3. И.В.Вельбицкий. Метаязык для формального задания семантики языков программирования. Препринт 75-78. Изд-во Института кибернетики АН УССР, Киев, 1975.
4. В.Н.Лебедев. Введение в системы программирования. Изд-во "Статистика", Москва, 1975.
5. А.С.Шумей, В.С.Зонис. О синтаксическом анализе по однозначным грамматикам. Программирование, № 3, 1975.
6. В.А.Сухомлин. Применение машины грамматического разбора в процессах трансляции. Программирование, № 3, 1975.
7. С.Я.Виленкин, С.М.Мовшорич. Повышение эффективности рабочих программ путем изменения исходной грамматики языка программирования. Программирование, № 1, 1975.

Графический способ задания синтаксиса команды формирования редакторского файла.

`CREATE [,line[,incr]] [,SUP]`

Квадратные скобки указывают на необязательность присутствия параметра. Подчеркнутые символы являются ключевыми для распознавания команды.

line - номер первой строки текста редакторского файла.

incr - шаг нумерации.

Рукопись поступила в издательский отдел
3 января 1978 года.