

**СТАТИСТИЧЕСКО ИЗСЛЕДВАНЕ НА АЛГОЛ-60****Маргарита Бърнева и Стоян Бъчваров**

Използването на алгоритмични езици, разработката на транслатори и създаването на нови алгоритмични езици налагат да се правят задълбочени изследвания върху структурата на съществуващите вече алгоритмични езици. При тези изследвания трябва да се имат пред вид както изискванията на потребителите (широки възможности на езиците, естествени ограничения и т. н.), така и възможността за създаване на прости и ефективни алгоритми за трансляция.

При създаването на транслатори е важно да се прави разумен компромис между избрания метод за трансляция и възможностите на изчислителната система, на която се реализира езикът. Биха могли да се реализират оптимални схеми само за онези категории на езика, които се срещат по-често. За рядко срещаните се категории алгоритмите за трансляция могат да бъдат по-прости, а реализиращите ги програмни части от транслатора могат да бъдат записани в памет с по-бавен достъп.

Един метод за изучаване на даден алгоритмичен език е да се изследват статистически по-характерните черти на програми, написани на него. Резултатите от такива изследвания могат да покажат нецелесъобразността на някои ограничения и да потвърдят други.

Известно е, че програмите се влияят силно от редица фактори, като естеството на задачите, които се решават, опитността и индивидуалните качества на програмиста и т. н. За да се направят правдоподобни изводи при статистическо изследване на програми, е необходимо да се обхване широк кръг програми.

В настоящата работа се разглежда езикът АЛГОЛ-60 и се определят характеристиките на езика, които представляват интерес за статистическо изследване. По разработените методи са създадени програми за автоматична сметачна машина (АСМ) „Минск-2“. Чрез тези програми са обработени всички процедури от [1], [2] и [3].

Въпреки ограничения обем на материала той е от такъв характер, че получените от обработката резултати позволяват да се установят известни особености и да се направят съответни изводи. Благоприятен фактор е, че процедурите са писани изобщо от опитни програмисти, които не са взимали пред вид ограничения, поставящи от конкретен транслатор. Фактът, че статистическите изследвания са извършени върху алгоритми,

записани на еталонния език АЛГОЛ-60 [4], позволява да се установят неизползваните (или рядко използвани) възможности на езика, както и разпределението и начините на използване на употребяваните възможности.

#### 1. НЕОБХОДИМОСТ ОТ СТАТИСТИЧЕСКО ИЗСЛЕДВАНЕ И ПО-ВАЖНИ ХАРАКТЕРИСТИКИ НА ПРОГРАМИ, НАПИСАНИ НА АЛГОЛ-60

Еталонният език АЛГОЛ-60 предоставя широки възможности за записване на алгоритми. В него не се правят никакви ограничения за броя на различните категории, включени в една програма. От друга страна, параметрите на АСМ, върху която се реализира езикът, налагат някои ограничения — например диапазона на използваните реални числа. Редица други ограничения се въвеждат при създаването на всеки транслятор. Те се дължат главно на ограничения обем на оперативната памет и на свързаното с това разпределение на паметта както по време на трансляцията, така и по време на изпълнение на програмата. За избягване на трудностите, възникващи от ограничеността на оперативната памет, се въвеждат например ограничения за броя и дължината на идентификаторите в дадена програма, за броя на целите и реалните числа, за дължината на блоковете, операторите, изразите и други елементи на програмата, за размерността на масивите, за дълбочината на вложеност на блоковете и т. н.

Поради липса на данни въвежданите ограничения често носят субективен характер и водят до неоптимално използване на ресурсите на АСМ. Така например дължината на идентификаторите в съкращения АЛГОЛ е най-много 6 символа, а във Вестоунския транслятор [6] — 8.

Ако в транслятора се използват отделни таблици на числата и идентификаторите, целесъобразно е съотношението на обемите на тези таблици да съответствува на средното статистическо съотношение между броя на числата и идентификаторите в програмата. По този начин ще се намали опасността едната от таблиците да се окаже препълнена, докато другата е полупразна. Така например в транслятора ТАМ-22 [5] се допуска използването на реални числа, цели числа и идентификатори в съотношение 1 : 11,6, докато в проведеното изследване на алгоритмите, публикувани в [1], [2] и [3], това съотношение е 1 : 16,8 : 12,4. Максималната допустима размерност на масивите в същия транслятор е 15, а максималното допустимо вложение на блокове — 36, докато в проведеното изследване на цитираните алгоритми се срещат масиви с размерност най-много 2 и дълбочина на вложение на блоковете най-много 4.

Разбира се, горните съпоставения не трябва да се абсолютизират, тъй като се отнасят, от една страна, за един не добре балансиран транслятор, а, от друга, за ограничен брой алгоритми от специален характер. Въпреки това приведените прости примери показват колко съществено е да се знаят реалните съотношения между различните елементи на програмите при проектирането на транслятори. Естествено освен средните стойности е интересно да се познават и съответните дисперсии, но познаването дори само на средните стойности позволява да се подобри съществено проектирането на даден транслятор.

С оглед на създаването на ефективни методи за трансляция е необходимо да се познават и някои други характеристики на потребителските програми като разпределението на операторите по типове, честотата на

срещането на различните служебни думи, дълбочината на вложение на цикли, характера на елементите на списъците на цикли, разпределението на формалните параметри на процедурите по спецификации, разпределението на идентификаторите по тип и вид, някои характерни разновидности на операторите, като броя на елементите на списъка на дясната част на операторите за присвояване, и т. н.

Познаването на горните характеристики на програмите позволява да се създадат оптимални по време схеми за трансляция, които да осигуряват бързо откриване, анализиране и обработка на често срещащите се елементи от програмите.

Извършването на статистически изследвания дава възможност да се решат обективно и редица спорни въпроси — необходимостта от рекурсивни процедури, от етикети, които са цели числа без знак, от собствени променливи и масиви и специално от собствени масиви с променливи граници, използването на променливи с индекси за параметри на цикли, именуващи изрази, съдържащи условия, и т. н.

Възможно е да се възрази, че средните статистически характеристики не са определящи, тъй като транслятор, съобразен с тях, няма да може да обработва програми, съдържащи изключителни случаи. Но такива програми ще се явяват рядко и това не е сериозно основание да се усложняват трансляторите и да се губи от тяхната ефективност, още повече че като правило езикът АЛГОЛ-60 позволява такива програми да се записват чрез често употребявани конструкции.

## 2. МЕТОД НА ПРОГРАМНА РЕАЛИЗАЦИЯ

Изследването на програми, написани на АЛГОЛ-60, се извършва чрез неколкостепенна обработка на текстовете. Програмите, написани на еталонен език, се кодират в конкретното представяне, използвано в транслятора ТАМ-22 [5], и се въвеждат във външната памет на машината. След това за всяка обработка цялата информация се въвежда на порции в оперативната памет. При обработването на информацията тя се преобразува в подходящо избран вътрешен код, съдържащ по 9 разряда за всеки основен символ на АЛГОЛ-60. Паралелно с преобразуването на информацията във вътрешен код се извършва контрол за допуснати грешки при кодирането в конкретното представяне и при перфорацията на данните.

При всеки преглед на информацията се определя зададена група статистически характеристики. Извършват се общо четири прегледа.

Програмната реализация е извършена в системата МИД-2 на машината „Минск-2“ на Математическия институт с Изчислителен център. Тя се състои от четири обработващи програми, съответстващи на четирите прегледа на информацията, и от управляваща програма, която ръководи четенето на информацията на порции, преобразуването ѝ във вътрешен код, натрупването и окончателната обработка на съответните резултати.

В обработващите програми се използват алгоритми, базирани на синтаксиса и семантиката на езика АЛГОЛ-60. Тези програми извършват синтаксически анализ, съответстващ на групата символи или понятия, които се изследват при дадена обработка и натрупват получените резултати.

Функциите на различните обработващи програми са следните.

Обработваща програма № 1 изследва разпределението на операторите, като отчита вложеността на блоковете и операторите за цикъл.

Обработваща програма № 2 изследва разпределението на идентификаторите по дължина и клас, относителната честота на простите променливи и масиви, разпределението на процедурите по тип, някои характеристики на променливите с индекси, разпределението на формалните параметри и използването на цяло без знак като етикет.

Програмата изследва само синтаксически единици, които се срещат в описанията. Паралелно с обработката се осъществява и известен синтаксически контрол и се дават индикации за грешки.

Обработваща програма № 3 изследва разпределението на служебните думи, някои характеристики на операторите за цикли, разпределението на числата, използването на ключови списъци, съдържащи само етикети.

Обработваща програма № 4 проверява в даден алгоритъм при последователни стойности на  $n$  колко различни идентификатора ще станат неразличими, ако се вземат под внимание само първите им  $n$  символа.

### 3. РЕЗУЛТАТИ И НЯКОИ ИЗВОДИ

Основната част от резултатите от обработката на всички процедури от [1], [2] и [3] са дадени в табл. 1, 2, 3, 4 и 5. Някои допълнителни резултати са изложени по-долу. Броят на процедурите от [1], [2] и [3] е 138, но тъй като някои от тях са в различни варианти или съдържат предварително описани помощни процедури, общият брой на всички обработени процедури е 165.

В табл. 1 е приведено процентното разпределение на служебните думи спрямо общия им брой 5691. Пред вид специфичността на обработ-

Таблица 1

Разпределение на служебните думи

Служебна дума	Относителна честота в %	Служебна дума	Относителна честота в %
<b>begin</b>	9,6	<b>array</b>	3,1
<b>end</b>	9,6	<b>value</b>	2,9
<b>if</b>	9,0	<b>else</b>	2,7
<b>then</b>	9,0	<b>comment</b>	1,5
<b>for</b>	8,0	<b>Boolean</b>	0,6
<b>do</b>	8,0	<b>true</b>	0,4
<b>step</b>	7,9	<b>false</b>	0,3
<b>until</b>	7,9	<b>label</b>	0,1
<b>integer</b>	5,5	<b>own</b>	0,1
<b>go to</b>	5,5	<b>while</b>	0,1
<b>real</b>	4,6	<b>switch</b>	0,1
<b>procedure</b>	3,5	<b>string</b>	0,0

вания материал процентът на **procedure**, **value**, а може би и **comment** е завишен. Ако не се вземат пред вид служебните думи **procedure** в началото на всеки от обработените алгоритми, процентното съдържание на

тази дума ще бъде около 0,9%. От друга страна, съотношението на служебните думи **procedure** и **value** показва, че в повече от 80% от процедурите съществуват формални параметри, които се заместват по стойност.

В табл. 2 се дава процентното разпределение на идентификаторите по дължина, т. е. по броя на участващите в тях символи, спрямо общия брой на идентификаторите 2131. От тази таблица личи, че повече от половината използвани идентификатори се състоят само от един символ; броят на идентификаторите с повече от 6 символа е 5,3%, а броят на идентификаторите с повече от 8 символа е само 2,4%.

Направеното изследване за броя на съвпадащите идентификатори в дадена процедура по първите си  $n$  символа дава следните резултати:

47,6% при  $n = 1$ ,  
 8,2% при  $n = 2$ ,  
 4,4% при  $n = 3$ ,  
 1,4% при  $n = 4$ ,  
 0,0% при  $n \geq 5$ .

Процентите са спрямо общия брой идентификатори. Тези резултати показват, че в обработените програми е достатъчно да се вземат под внимание само първите 5 символа на идентификаторите.

В табл. 3 е приведено разпределението на идентификаторите по тип и вид спрямо общия им брой (2131). От общия процент (8,4%) на идентификаторите за процедури 8,2% са идентификатори на процедури с параметри и 0,2% — на процедури без параметри.

По-голямата част (около 75%) от идентификаторите на процедури не са определени по тип, което показва, че около 25% от всички процедури са процедури-функции. Това съотношение е характерно за обработвания материал. Специфичен за обработвания материал е и значителният процент формални параметри, идентификатори на процедури и частично — етикети. Процентът на идентификатори на ключове е незначителен.

Таблица 2

Разпределение на идентификаторите по дължина

Дължина на идентификатора	Относителна честота в %
1 символ	54,0
2 символа	15,7
3 символа	9,3
4 символа	7,7
5 символа	5,8
6 символа	2,2
7 символа	1,9
8 символа	1,0
9 символа	0,7
10 символа	0,5
11 символа	0,4
12 символа	0,4
13 символа	0,1
Повече от 13 символа	0,3

Таблица 3

Разпределение на идентификаторите по тип и вид

Вид на идентификатора	Относителна честота в %			
	real	integer	Boolean	общо
Проста променлива	18,5	16,6	0,6	35,7
Променлива с индекс	4,8	1,3	0,4	6,5
Идентификатор на процедура	1,8	0,1	0,1	8,4
Формални параметри	—	—	—	36,8
Етикети	—	—	—	12,5
Идентификатори на ключове	—	—	—	0,1

Направеното изследване за броя на индексите в променливите с индекс показва, че в 88% от случаите променливите съдържат по един индекс, а в останалите 12% — два индекса. Променливи с повече от два индекса не се използват.

Във връзка с разпределението по типове може да се отбележи, че общият брой (3067) константи, използвани в операторите, се разпределя така: 4,2% тип реален, 94,5% тип цял и 1,3% тип Булев, т. е. почти изключително се използват константи от тип цял. При това от използваните 2895 цели числа в 17,5% от случаите числото е нула, в 67,4% — числото е 1 и в 15,1% — числото е 2.

Таблица 4  
Разпределение на формалните параметри по спецификации

Спецификация на формалния параметър	Относителна честота		
	извикан по име		общо
real	14,8	20,8	35,6
integer	6,2	22,7	28,9
Boolean	0,6	0,5	1,1
real procedure	2,5	0,0	2,5
integer procedure	0,1	0,0	0,1
Boolean procedure	0,3	0,0	0,3
procedure	0,7	—	0,7
real array	0,0	—	0,0
integer array	4,9	—	4,9
Boolean array	0,9	—	0,9
array	22,1	—	22,1
switch	0,0	—	2,0
string	0,0	—	0,0
label	0,9	—	0,9
Неспецифицирани	2,0	—	2,0
Всичко	56,0	44,0	100,0

В табл. 4 се дава процентното съдържание на формалните параметри на процедурите по спецификации спрямо общия им брой 792. Както се вижда от таблицата, основните видове спецификации са **real**, **integer** и **array** — общо 86,6%; параметрите, замествани по стойност, са 44% спрямо всички параметри, но 64% спрямо параметрите, за които формално е допустимо да бъдат заместени по стойност. Прави впечатление слабо използване на спецификациите **string**, **switch**, **label** и **procedure**.

В табл. 5 е приведено процентното разпределение на операторите за преход, за процедура, за цикъл, на условните и съставните оператори и блоковете спрямо операторите за присвояване. Като база е избран броят на операторите за присвояване, а не общият брой на всички оператори, тъй като не е естествено да се събират другите типове оператори със съставните оператори и блоковете, които вътре в себе си съдържат други оператори.

Общият брой на операторите за присвояване е 1977. Болшинството от тях (94,8%) са прости; в 4,2% списъкът на лявата част съдържа 2 променливи, в 0,7% списъкът на лявата част съдържа 3 променливи, а списъци с 4, 5 и 7 променливи се срещат по в 0,1% от общия брой на операторите за присвояване.

Таблица 5  
Разпределение на операторите спрямо операторите за присвояване

Тип на оператора	Процент спрямо операторите за присвояване	
За преход	15,9	
За процедура	5,1	
Условен	21,4	
в това число :		
условен — непълнен		18,6
условен — пълнен		2,8
За цикъл	23,1	
в това число :		
първо ниво		15,7
второ ниво		5,8
трето ниво		1,5
четвърто ниво		0,1
Съставен	20,1	
Блок	8,2	
в това число :		
първо ниво		7,1
второ ниво		0,9
трето ниво		0,1
четвърто ниво		0,1
Общо	93,8	

От табл. 5 личи, че общият брой на другите типове оператори е почти равен на броя на операторите за присвояване; един съставен оператор се пада средно на 5 оператора за присвояване, а един блок съдържа средно 12 оператора за присвояване. Значителната честота на блоковете се дължи на сравнително кратките програми в обработвания материал.

Както се вижда от табл. 5, практически се срещат само прости (до второ ниво) вложения на блокове. Вложения от трето и четвърто ниво се срещат твърде рядко.

При циклите също преобладават простите цикли (първо ниво), но се срещат често и прости вложения на цикли.

Прави впечатление рядкото използване на пълния условен оператор (if-then-else) — само в 13% от случаите на използване на условни оператори.

По отношение вида на елементите на списъците на операторите за цикли се получават следните резултати:

- 3,8% са аритметичен елемент,
- 1,1% итеративен елемент (*A while B*),

95,1% стъпков елемент (*A1 step A2 until A3*), в това число елементи от вида: *A2 1, A3* = число или проста променлива — 77,3%, *A2 1, A3* число или проста променлива — 14,0%, *A2* = число, различно от 1 или проста променлива, *A3* — число или проста променлива — 8,7%.

Параметрите на циклите са почти винаги прости променливи. Само в 0,9% от случаите се срещат променливи с индекси, използвани за параметри на цикли.

Сред използваните етикети не се срещат етикети от вида цяло без знак.

Приведените резултати могат да дадат повод за редица съпоставяния и да позволят да се направят известни заключения както за характера на езика АЛГОЛ-60 и въвежданите ограничения при различни реализации, така и за по-ефективна организация на трансляторите.

#### ЛИТЕРАТУРА

1. Агеев, М. И., В. П. Алик, Р. М. Галис. Алгоритмы (1—50). М., 1966.
2. Агеев, М. И., В. П. Алик, Л. В. Малюк, Ю. И. Марков. Алгоритмы (51—100). М., 1966.
3. Агеев, М. И., Л. С. Кривонос, Ю. И. Марков. Алгоритмы (101—150), М., 1967.
4. Алгоритмический язык АЛГОЛ-60, пересмотренное сообщение. М., 1965.
5. Демидович, Н. В., Г. С. Шинкевич, Н. В. Шкут. Трансляторы ТАМ-2 и ТАМ-22. Минск, 1967.
6. Ренделл, Б., Л. Рассел. Реализация АЛГОЛА-60. М., 1967.

*Постъпила на 9. I. 1970 г.*

## СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ АЛГОЛ-60

Маргарита Бырнева и Стоян Бычваров

*(Резюме)*

Рассматривается необходимость глубокого изучения характерных особенностей программ, написанных на АЛГОЛ-60, в связи с введением ограничений в конкретных реализациях и подбором оптимальных алгоритмов для трансляции.

Рассматривается ряд статистических характеристик программ, написанных на АЛГОЛ-60, приводится описание метода и соответствующей программной реализации, по которой возможно определение этих характеристик.

Приведены результаты проведенных статистических исследований процедур из [1], [2] и [3]. Основные результаты даны в табл. 1, 2, 3, 4 и 5.



# A STATISTICAL STUDY OF ALGOL-60

Margarita Bârneva and Stojan Báčvarov

*(Summary)*

The necessity for a profound study of the characteristic features of programmes written in ALGOL-60 is considered in order to introduce some restrictions in the concrete realizations and to choose the optimal algorithms for translation.

A number of statistical characteristics of programmes written in ALGOL-60 are examined and a method and a concrete programme realization enabling us to determine the characteristics under discussion are given.

The results from the statistical analysis of the procedures in [1], [2] and [3] are given. The tables 1, 2, 3, 4, 5 contain the basic results obtained.