

УДК 62.506.2

М. Ф. БОНДАРЕНКО, канд. техн. наук, А. Ф. ОСЫКА, Н. К. СВИНАРЬ,
А. И. ЧУГУН

АЛГОРИТМ ПЕРЕВОДА КОЛИЧЕСТВЕННЫХ И ПОРЯДКОВЫХ ЧИСЛИТЕЛЬНЫХ В ЦИФРОВУЮ ЗАПИСЬ

Описанный в данном сообщении алгоритм моделирует одну из функций человеческого мозга, связанную с анализом и переработкой языковой информации. На вход описываемого алгоритма могут поступать простые, сложные и составные количественные и порядковые числительные в любом падеже, написанные в соответствии с правилами орфографии русского языка. На выходе алгоритма получается цифровой эквивалент исходного числительного с соответствующим окончанием. Например, восемьдесятю одним = 81-м.

Иными словами, на вход описываемой модели поступает последовательность $X_1, X_2, \dots, X_i, \dots, X_n$ ($n=1, \dots, 18$), где X_i представляет собой определенную падежную форму числительного, которое пишется одним словом. Если на вход поступает простое или сложное числительное, то $n=1$. Если же на входе составное числительное, то $n>1$.

Постоянная информация, помещенная в таблице, позволяет анализировать входные составные числительные, состоящие не более чем из 23 слов. Это обусловлено тем, что в таблицу внесены основы числительных, включая квадрильон. Увеличение максимальной длины исходной последовательности можно было бы достигнуть за счет включения в таблицу названий чисел больше, чем квадрильон. Эти названия, однако, практически не употребляются, хотя и включены в орфографический словарь [1].

Такие большие числа не записываются словами или цифрами, а фиксируются в виде степеней числа 10. В случае расширения

<i>j</i>	В	Ц	П
1	ДЕСЯТ	1	1
2	ДЕВЯНОСТ	9	1
3	СОРОК	4	1
4	СТ	1	2
5	ТЫСЯЧ	1	3
6	МИЛЛИОН	1	6
7	МИЛЛИАРД	1	9
8	ТРИЛЛИОН	1	12
9	КВАДРИЛЬОН	1	15
10	ДВАДЦАТ	2	1
11	ТРИДЦАТ	3	1
12	ДЕВЯТ	9	0
13	ВОСЕМ	8	0
14	ВОСЬМ	8	0
15	СЕМ	7	0
16	ШЕСТ	6	0
17	ПЯТ	5	0
18	ЧЕТЫР	4	0
19	ТР	3	0
20	ДВ	2	0
21	ОДИН	1	0
22	ОДН	1	0
23	НАДЦАТ	1	1
24	ДЕСЯТ	1	1
25	СОТ	1	2
26	СТ	1	2
27	СОРОКОВ	4	1
28	ТРЕТ	3	0
29	ЧЕТВЕРТ	4	0
30	СЕДЬМ	7	0
31	ВТОР	2	0
32	ПЕРВ	1	0
33	СОТ	1	2

таблицы постоянной информации за счет включения в нее редких названий больших чисел (квинтильон, секстильон и т. д.) предлагаемый ниже алгоритм изменений не потребует. Изменится лишь соответствующим образом область допустимых входных сигналов. На этом основании данный алгоритм можно считать кибернетической моделью одной из функций человеческого мозга, связанной с анализом числительных русского языка.

Предлагаемый алгоритм отличается от описанного в работе [2] алгоритма перевода названий целых чисел следующим:

1) алгоритм [2] работает лишь с количественными числительными. Предлагаемый алгоритм работает на множестве количественных и порядковых числительных;

2) алгоритм [2] работает лишь с числительными в именительном падеже. Предлагаемый алгоритм преобразует числительные, стоящие в любом падеже;

3) алгоритм [2] рассчитан на анализ слов, обозначающих название чисел до миллиона, т. е. может работать, хотя и на очень большом, но все же ограниченном множестве входных сигналов

Предлагаемый алгоритм работает на множестве входных сигналов, которое практически не ограничено;

4) алгоритм [2] переводит числительное не в цифровую запись непосредственно, а в запись на языке-посреднике. Например, двадцать семь=2, 1; 7, 0. Предлагаемый алгоритм переводит числительное непосредственно в цифровую запись с соответствующим окончанием. Например, двадцати семи=27-и;

5) алгоритм [2] рассчитан на анализ слов, обозначающих числа, которые состоят не более чем из двух корней (шестьдесят восемьдесят и т. д.). Предлагаемый алгоритм анализирует сложное числительное, состоящее из любого количества корней (на пример, шестисотвосемьдесятпяти миллионный), так как числительные, оканчивающиеся на -тысячный, -миллионный и т. д., могут иметь в своем составе до шести корней [3].

Необходимая для работы описанного ниже алгоритма постоянная информация помещена в таблице. В графе *j* таблицы помещены номера основ числительных. В графе В помещены основы числительных. В графе Ц помещены значащие цифры числовы

эквивалентов основ. В графе П указано количество нулей, которое следует приписать к значащей цифре, чтобы получить числовой эквивалент соответствующей основы.

Для работы алгоритма используются два условных элемента памяти: М и К.

Описание алгоритма

1. Взять очередную (начиная с первой) словоформу исходной последовательности и проверить, является ли она последней на входе. Если да, то перейти к п. 2, иначе — к п. 3.

2. Проверить, совпадает ли начало анализируемой словоформы с одной из основ ($27 \leq j \leq 33$) таблицы. В случае положительного ответа перейти к п. 4, иначе — к п. 3.

3. Найти в таблице первую основу ($1 \leq j \leq 22$), которая совпадает с началом анализируемой словоформы. Перейти к п. 4.

4. Вычеркнуть из анализируемой словоформы совпадающую часть и предшествующие ей буквы (если они имеются). Перейти к п. 5.

5. Проверить, является ли содержание графы П у последней выделенной основы больше 2. Если да, то перейти к п. 6, в противном случае перейти к п. 14.

6. Проверить, есть ли записи в элементе памяти М. Если да, то перейти к п. 7, если нет — к п. 13.

7. Содержимое элемента памяти К увеличить на число, равное $M \cdot 10^{P_j}$ (где М — содержимое элемента памяти М; P_j — содержимое графы П у последней выделенной основы). Стереть содержимое элемента памяти М. Перейти к п. 8.

8. Проверить, является ли анализируемая словоформа последней в исходной последовательности. Если да, то перейти к п. 9, иначе — к п. 1.

9. Содержимое элемента памяти К увеличить на число, равное содержимому элемента памяти М. Перейти к п. 10.

10. Проверить, остались ли буквы в последней словоформе после удаления из нее выделенных основ. Если да, то перейти к п. 11, иначе — к п. 12.

11. Записать на выходе: $X_1, X_2, \dots, X_i, \dots, X_n$ — содержимое элемента памяти К, тире, остаток последней анализируемой словоформы. Перейти к п. 24.

12. Записать на выходе: $X_1, X_2, \dots, X_i, \dots, X_n$ — содержимое элемента памяти К. Перейти к п. 24.

13. Содержимое элемента памяти М увеличить на число, равное $\Pi_j \cdot 10^{P_j}$ (где Π_j и P_j — содержимое граф соответственно Π и P у последней выделенной основы). Перейти к п. 8.

14. Проверить, является ли номер последней выделенной основы больше 26. Если да, то перейти к п. 13, иначе — к п. 15.

15. Проверить, содержит ли анализируемая словоформа после удаления выделенных основ больше трех букв. Если да, то перейти к п. 16, если нет — к п. 13.

16. Проверить, совпадает ли одна из основ ($23 \leq j \leq 26$) таблицы с остатком анализируемой словоформы или его частью (совпадение должно произойти не обязательно, начиная с первой буквы остатка, но не позднее четвертой). Если да, то перейти к п. 17, иначе — к п. 23.

17. Удалить из остатка анализируемой словоформы совпадающую часть и предшествующие ей буквы (если они имеются). Перейти к п. 18.

18. Проверить, совпадает ли последняя выделенная основа с *надцат*. Если совпадает, то перейти к п. 19, если нет — к п. 20.

19. Содержимое элемента памяти М увеличить на число, равное $\Pi_{j-1} + 10$ (где Π_{j-1} — содержимое графы Π у предпоследней выделенной основы). Перейти к п. 21.

20. Содержимое элемента памяти М увеличить на число, равное $\Pi_{j-1} \cdot 10^{P_j}$ (где Π_{j-1} — содержимое графы Π у предпоследней выделенной основы, P_j — содержимое графы P у последней выделенной основы). Перейти к п. 21.

21. Проверить остаток анализируемой основы. Если он содержит больше трех букв, то перейти к п. 22, в противном случае перейти к п. 8.

2. Найти в таблице первую основу ($1 \leq j \leq 22$), которая совпадает с остатком анализируемой словоформы или его частью (совпадение должно произойти не обязательно начиная с первой буквы остатка, но не позднее четвертой). Перейти к п. 4.

23. Содержимое элемента памяти M увеличить на число, равное $C_j \cdot 10^{P_j}$ (где C_j и P_j — содержимое граф соответственно C и P у последней выделенной основы). Перейти к п. 22.

24. Конец работы алгоритма.

Данный алгоритм положен в основу программы для ЭВМ «Минск-32» и может быть использован для решения задач по анализу и обработке словесной информации.

СПИСОК ЛИТЕРАТУРЫ

1. Орфографический словарь русского языка. М., «Сов. энциклопедия», 1971. 520 с.
2. Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста. Вып. 3. М., изд. АН СССР, 1961. 12 с.
3. Супрун А. Е. О русских числительных. Фрунзе, Изд-во Киргизск. ун-та, 1959. 140 с.
4. Грамматика русского языка. Т. 1. М., изд. АН СССР, 1960. 720 с.

Поступила 26 ноября 1974 г