

## ОБ ОДНОМ АЛГОРИТМЕ СИНТЕЗА ЧИСЛИТЕЛЬНЫХ РУССКОГО ЯЗЫКА

Рассмотрим укрупненную блок-схему алгоритма, который моделирует работу человеческого мозга при синтезе количественных и порядковых числительных русского языка. На вход алгоритма поступают произвольные цифровые последовательности с соответствующими окончаниями любого падежа, которые могут указывать на количественные либо порядковые числительные. В качестве частного случая возможно нулевое окончание у цифровой последовательности на входе, которая в дальнейшем будет именоваться входной. В этом случае алгоритм выдает соответствующее количественное числительное в именительном падеже. На выходе алгоритм вырабатывает количественное или порядковое числительное, соответствующее входной последовательности и стоящее в падеже, которого требует ее буквенное окончание. Например, входной последовательности *243-х* алгоритм поставит в соответствие выходную последовательность *двадцать сорока трех*.

В работе [1] предложены частные алгоритмы синтеза количественных числительных русского языка в именительном падеже, закодированных на специальном языке-посреднике. Алгоритм, предлагаемый в настоящем сообщении, синтезирует количественные и порядковые числительные в произвольном падеже.

Алгоритм [1] может синтезировать сложные числительные, состоящие не более чем из двух основ (восемьсот, шестьдесят и т. д.). Описываемый нами алгоритм в случае необходимости может синтезировать сложные числительные, состоящие более чем из двух основ (шестисотсемидесятипяти тысяч, двадцатидвухмиллионный и т. д.).

Алгоритм [1] может синтезировать названия чисел от 1 до 999 999, т. е. работает хотя и на очень большом, но все же ограниченном множестве входных сигналов. Предлагаемый в данной работе алгоритм может синтезировать числительные, соответствующие как угодно большим числам, для которых существуют названия в русском языке, т. е. работает на неограниченном множестве входных сигналов.

Построение алгоритма синтеза числительных сводится к пяти основным этапам:

1) изучение закономерностей сочетаемости составных элементов числительных;

2) определение достаточного набора формальных признаков для синтеза количественных и порядковых числительных;

3) выявление основных этапов синтеза числительных и разработка подробного алгоритма;

4) психологический эксперимент в целях проверки соответствия результатов работы алгоритма данным носителя русского языка;

5) написание программы и проверка на ЭВМ модели синтеза числительных.

Необходимая для работы описываемого алгоритма постоянная информация (составные элементы числительных и их формальные признаки) включена в две таблицы. В первой приводятся основы количественных числительных и варианты этих основ (*один, одн, восемь, восьм* и т. д.), а также специфические основы порядковых числительных (*перв, втор* и т. д.). Каждой основе в этой таблице приписана следующая информация:

а) типы склонений количественных и порядковых числительных, с окончаниями которых сочетается данная основа;

б) стоящие в постпозиции в составе сложного числительного основы, с которыми сочетается данная основа;

в) значащая цифра цифрового эквивалента данной основы;

г) число нулей, которое следует приписать справа к значащей цифре, чтобы получить цифровой эквивалент данной основы.

Во второй таблице помещены окончания количественных и порядковых числительных, сгруппированные по типам склонений. По номеру строки в этой таблице можно установить, принадлежит ли данное окончание количественному или порядковому числительному, а по номеру графы — падеж окончания.

Алгоритм обрабатывает входную последовательность справа налево, разбивая ее в процессе работы на группы по три цифры — триады. Опишем укрупненную блок-схему, которая указывает на основные этапы работы алгоритма при синтезе числительных.

Блок 1. Выбор из первой таблицы основы числительного (или двух основ), соответствующей очередной анализируемой цифре входной последовательности. Переход к блоку 2.

Блок 2. Проверка: определен ли уже падеж синтезируемого числительного. Если «да», то переход к блоку 6, «нет» — к блоку 3.

Блок 3. Сравнение буквенного окончания входной последовательности с окончаниями во второй таблице тех типов склонения, которые указаны в первой таблице для основы, взятой в блоке 1. Переход к блоку 4.

Блок 4. Выбор из второй таблицы окончания (или двух окончаний), соответствующего падежу синтезируемого числительного. Формирование и запись очередного слова, входящего в состав синтезируемого числительного и соответствующего анализируемой цифре входной последовательности. Переход к блоку 5.

Блок 5. Проверка: все ли цифры входной последовательности проанализированы. Если «да», то переход к блоку 8, «нет» — к блоку 1.

Блок 6. Проверка: является ли последняя основа, полученная в результате работы блока 1, основой *тысяч, миллион, миллиард* и т. д. Если «да», то перейти к блоку 7, «нет» — к блоку 4.

Блок 7. Выбор из второй таблицы окончания, соответствующего последней взятой из первой таблицы основе. Число и падеж окончания зависят не только от падежа всего синтезируемого числительного, но и от цифры, которая стоит первой справа в анализируемой триаде. Формирование и запись очередного слова, входящего в состав синтезируемого числительного. Переход к блоку 5.

Блок 8. Замена порядка следования слов, входящих в состав синтезируемого числительного, на противоположный (это делается потому, что цифры входной последовательности анализируются справа налево).

Конец работы алгоритма.

Поясним содержание работы некоторых блоков. В блоке 1 анализируемой цифре входной последовательности ставится в соответствие только одна основа, если в первой таблице имеется основа, у которой в информации в пункте в) стоит число, равное анализируемой цифре входной последовательности, а в пункте г) записано число, на единицу меньше порядкового номера анализируемой цифры в триаде (при отсчете справа налево). Если информация в г) не соответствует номеру анализируемой цифры, то берется одна из основ, записанная в пункте б), у которой количество нулей цифрового эквивалента на единицу меньше номера анализируемой цифры в триаде. В этом случае анализируемой цифре ставятся в соответствие две основы. Основы *тысяч, миллион, миллиард* и т. д. выбираются перед началом анализа очередной триады.

В блоке 2 ответ «нет» вырабатывается только при обработке первой значащей (отличной от нуля) цифры входной последовательности, если на входе имеется буквенное окончание. В остальных случаях падеж всего синтезируемого числительного и каждого отдельного слова, входящего в его состав, считается известным.

Если в блоке 1 взята одна основа, то в блоке 4 выбирается одно необходимое окончание из тех типов склонения, которые записаны в информации к основе в пункте а). Если в блоке 1 взято две основы, то выбираются два соответствующих этим основам окончания. При формировании очередного выходного слова учитывается то обстоятельство, что, как правило, одной значащей (отличной от нуля) цифре входной последовательности соответствует одно слово, входящее в состав выходного числительного. Исключение составляют случаи, когда на выходе синтезируется слово с основой *надцат*, а также заканчивающееся на *-тысячный, -миллионный* и т. д. [2].

Число и падеж слов с основами *тысяч, миллион, миллиард* и т. п. зависят не только от падежа синтезируемого числитель-

ного, но и от первой справа цифры в триаде. На это указывают следующие примеры: 1000 — одна тысяча; 2000 — две тысячи; 5000 — пять тысяч; 1000-е — одной тысяче; 2000-м — двум тысячам; 5000-м — пяти тысячам. Подобная зависимость учитывается при работе блока 7.

Описанная блок-схема была положена в основу подробного алгоритма, реализованного на ЭВМ «Минск-32». Созданная модель синтеза количественных и порядковых числительных русского языка может представлять самостоятельный интерес как математическое описание одной из психических функций мозга человека, а также найти применение при автоматическом переводе, в автоматизированных справочно-информационных системах и в ряде других случаев.

#### ЛИТЕРАТУРА

1. Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста. Вып. 3. М., Изд-во АН СССР, 1961. 12 с.
2. Розенталь Д. Э. Русский язык. Пособие для поступающих в вузы. М., Изд-во Моск. ун-та, 1965. 240 с.
3. Грамматика русского языка. Т. 1. М., Изд-во АН СССР, 1960. 719 с.