

МАШИННИЙ АЛГОРИТМ ПЕРЕНОСУ СЛІВ ДЛЯ УКРАЇНСЬКОЇ МОВИ

Для одержання за допомогою ЕОМ повнокодової перфострічки — програми керування рядковідливним складальним автоматом необхідно для рядків, які не можуть бути закінчені без переносу, здійснити автоматичний перенос частини слова з одного рядка в інший.

Зі всіх існуючих методів розділення слів на склади для української мови прийнятий логічний метод. За цим методом розділення слів на склади ведеться шляхом аналізу приголосних і голосних букв слова. Алгоритми поділу слів на склади розроблялись на основі «Українського правопису».

Нижче наведено алгоритми переносу слів для української мови.

1. Визначається кількість голосних і приголосних у слові. Слово, яке містить менше двох голосних або приголосних, перенести не можна.

2. Якщо слово містить більше чотирьох букв, його порівнюють зі словником префіксів для визначення наявності в ньому префікса. Коли слово містить префікс — перехід до п. 3, інакше — до п. 4.

3. Після префікса можливий перенос, крім випадків, вказаних у примітці до п. 3. Частина слова, що залишається після відокремлення префікса, аналізується на переносимість як слово, що не містить префікса.

Примітка. Якщо префікс закінчується на приголосну (крім букви «х»), а далі йде голосна, перенос можливий як перед цією приголосною, так і після неї.

4. Частина слова, що не містять голосних або приголосних, переносити не можна.

5. Перенос можливий після голосної перед приголосною.

6. Дві однакові приголосні, що знаходяться між голосними, розриваються переносом. Перед ними перенос заборонений. Це не накладає заборон на решту можливих переносів у слові.

7. При сполученні декількох приголосних перенос можливий у будь-якому місці між ними. Перед першою і останньою приголосними перенос заборонений.

Примітка. Буквосполучення «дж», «дз» і «ст» не можна розривати переносом, крім випадку, коли приголосні «д» або «с» відносяться до префікса. Тоді ці буквосполучення можна розривати.

8. Перед знаком «'» (апостроф) і після нього перенос заборонений, перед знаками «ь» і «й» перенос заборонений, після них можливий, крім випадків, даних у примітці.

Примітка. Буквосполучення «йю», «ьо» розривати не можна, причому перед «йю» перенос можливий.

9. При сполученні декількох голосних їх можна розділити переносом. Перед першою і після останньої перенос заборонений.

10. Після знаків дефіс і тире перенос можливий. Перед ними — заборонений.

Примітка. Частини слова, розміщені по різні боки від дефіса, аналізуються на переносимість як самостійні частини.

Для зменшення помилок, що виникнуть при використанні цих алгоритмів переносу, у пам'ять ЕОМ вводиться масив префіксів і найбільш часто вживаних перших частин складних слів. Використання словника префіксів значно збільшує точність поділу слів на склади.

У загальній задачі розрахунку виключки та формування рядка, власне машинні алгоритми переносу слів розділені на три частини.

1. Пошук частини слова, в якій треба шукати переноси.
2. Пошук префіксів та їх переносів.
3. Пошук переносів слова.

У першій частині алгоритму від слова відкидається частина, що не має голосної або приголосної, оскільки вона програмою пошуку переносів не буде відриватися від слова дефісом. Якщо слово закінчується на «й», «ь», то ці букви відкидаються і ведеться пошук до місця першої зміни голосної на приголосну або навпаки.

У другій частині порівнюється початок переносованого слова з префіксами з масиву префіксів. Якщо такий префікс є, то його переноси запам'ятовуються, а решта частини слова аналізується на перенос, починаючи з першої букви після префікса. Якщо префікса немає, ціле слово йде на блок переносів.

У третій частині одержуємо масив можливих переносів. Кількість їх також фіксується.

Для полегшення розуміння алгоритмів дамо деякі ознаки, які розміщені в розрядній сітці комірки оперативної пам'яті ЕОМ.

2-й розряд	=	1	—	буква
11-й	"		—	голосна
12-й	"		—	приголосна
13-й	"		—	й, ь

Далі під словом будемо розуміти сукупність букв a_1, a_2, \dots, a_n , де n — кількість букв у слові.

Для першої частини виборка букв починається з кінця слова.

Машинний алгоритм

Частина 1

1. Виконати: $i = n - 1$.
 2. Виконати: 11, 12 розряди $a_{i+1} \Rightarrow R_2$, де R_2 — робоча комірка.
 3. Перевірити: $R_2 = 0$?
- Якщо умова виконана, перейти до п. 4, якщо ні — до п. 5.
4. Виконати: 11, 12 розряди $a_i \Rightarrow R_2$.
 5. Перевірити: $i \geq 0$?
- Якщо умова виконана, перейти до п. 8, якщо ні — до п. 6.

6. Виконати: $0 \Rightarrow$ ДПС, де ДПС — довжина слова, що переноситься.
- Тут і далі для скорочення позначень використовуватимемо перші букви або склади повної назви.
7. Процес закінчити.
8. Виконати: $11, 12$ р. $a_{-1} i \Rightarrow R_1$, де R_1 — робоча комірка.
9. Перевірити: $R_1 = 0$?
- Якщо умова виконана, перейти до п. 5, якщо ні — до п. 10.
10. Перевірити: $R_1 = R_0$?
- Якщо умова виконана, перейти до п. 5, якщо ні — до п. 11.
11. Виконати: $i + 1 \Rightarrow$ ДПС.
12. Процес закінчити.

Частина 2

- Тут слово: a_1, a_2, \dots, a_k , де $k < n$.
1. Вибрати a_1 .
 2. По ходу a_1 визначити кількість префіксів на дану букву (КІЛПР).
 3. Перевірити: $КІЛПР = 0$?
 - Якщо умова виконана, перейти до п. 25, якщо ні — до п. 4.
 4. Визначити початкову адресу масиву префіксів на дану букву.
 5. Виконати: кількість букв префікса на дану букву \Rightarrow КІЛБП.
 6. Перевірити: $ДПС - КІЛБП \geq 0$?
 - Якщо умова виконана, перейти до п. 7, якщо ні — до п. 23.
 7. Перевірити: $5 - КІЛБП \geq 0$?
 - Якщо умова виконана, перейти до п. 9, якщо ні — до п. 8.
 8. Організувати лічильник для порівняння початку слова з префіксом. (При $КІЛБП > 5$).
 9. Організувати лічильник для порівняння початку слова з префіксом. (При $КІЛБП \leq 5$).
 10. Перевірити: префікс збігається з початком слова?
 - Якщо умова виконана, перейти до п. 11, якщо ні — до п. 23.
 11. Перевірити: $5 - КІЛБП < 0$?
 - Якщо умова виконана, перейти до п. 12, якщо ні — до п. 13.
 12. Перевірити: початок слова, починаючи з шостої букви, порівнюється з рештою частиною префікса?
 - Якщо умова виконана, перейти до п. 13, якщо ні — до п. 23.
 13. Перевірити: $ДПС - КІЛБП > 0$?
 - Якщо умова виконана, перейти до п. 14, якщо ні — до п. 25.
 14. Перевірити: $ДПС - КІЛБП = 0$?
 - Якщо умова виконана, перейти до п. 25, якщо ні — до п. 15.
 15. Визначити адресу масиву переносів для даного префікса.
 16. Перевірити: $КІЛПП = 0$? $КІЛПП$ — кількість переносів префікса.
 - Якщо умова виконана, перейти до п. 21, якщо ні — до п. 17.
 17. Перевірити: знак комірки $КІЛПП$ рівний «+»?
 - Якщо умова виконана, перейти до п. 19, якщо ні — до п. 18.
 18. Перевірити: 11 р. $a_{r+1} \neq 0$?, де r — кількість букв префікса.
 - Якщо умова виконана, перейти до п. 25, якщо ні — до п. 19.
 19. Виконати: $КІЛПП \Rightarrow$ комірку ЗВІТ. Масив переносів префікса \Rightarrow П 1, де П 1 — початкова адреса масиву переносів слова.
 20. Перевірити: a_{r+1} не апостроф?
 - Якщо умова виконана, перейти до п. 25, якщо ні — до п. 21.
 21. Виконати: $КІЛБП \Rightarrow$ РО.
 - $ДПС - КІЛБП - 1 \Rightarrow$ Р 1, де РО, Р 1 — робочі комірки.
 22. Перейти до п. 26.
 23. Перейти до наступного префікса.
 24. Перевірити: закінчився масив префіксів на дану букву?
 - Якщо умова виконана, перейти до п. 25, якщо ні — до п. 5.
 25. Виконати: $0 \Rightarrow$ РО.
 - $ДПС - 1 \Rightarrow$ Р 1.
 26. Процес закінчити.

Частина 3

Слово: a_1, a_2, \dots, a_k , де $k \leq p < n$: p — довжина слова після роботи першої частини.

1. Виконати: $0 \Rightarrow P+5$, де $P+5$ — робоча комірка дозволу переносу.
 2. Виконати: 11, 12 р. $a_k \Rightarrow R_2$ (тут $k=1$).
 3. Перевірити: $R_2=0$?
- Якщо умова виконана, перейти до п. 4, якщо ні — до п. 5.
4. Виконати: 11, 12 р. $a_k \Rightarrow R_2$ ($k=2$).
 5. Виконати: $k+1 \Rightarrow k$.
 6. Перевірити: $k \leq p$?
- Якщо умова виконана, перейти до п. 8, якщо ні — до п. 7.
7. Перейти до п. 49.
 8. Виконати: 11, 12, 13 р. $a_k \Rightarrow R_1$.
 9. Перевірити: $R_1=0$?
- Якщо умова виконана, перейти до п. 5, якщо ні — до п. 10.
10. Перевірити: $k_1=R_2$?
- Якщо умова виконана, перейти до п. 5, якщо ні — до п. 11.
11. Виконати: $R_1 \Rightarrow R_2$.
 12. Виконати: $a_k \Rightarrow R_4$, де R_4 — робоча комірка.
 13. Виконати: $k+1 \Rightarrow k$.
 14. Перевірити: $k \leq n$?
- Якщо умова виконана, перейти до п. 49, якщо ні — до п. 15.
15. Виконати: $a_k \Rightarrow R_3$, де R_3 — робоча комірка.
 16. Виконати: 11, 12, 13 р. $a_k \Rightarrow R_1$.
 17. Перевірити: R_4 — це апостроф?
- Якщо умова виконана, перейти до п. 41, якщо ні — до п. 18.
18. Перевірити: R_3 — це апостроф?
- Якщо умова виконана, перейти до п. 41, якщо ні — до п. 19.
19. Перевірити: 11 р. $R_4=0$?
- Якщо умова виконана, перейти до п. 20, якщо ні — до п. 25.
20. Перевірити: R_4 — це буква «д»?
- Якщо умова виконана, перейти до п. 21, якщо ні — до п. 22.
21. Перевірити: R_3 — це буква «ж»?
- Якщо умова виконана, перейти до п. 41, якщо ні — до п. 24.
22. Перевірити: R_4 — це буква «с»?
- Якщо умова виконана, перейти до п. 23, якщо ні — до п. 25.
23. Перевірити: R_3 — це буква «т»?
- Якщо умова виконана, перейти до п. 41, якщо ні — до п. 25.
24. Перевірити: R_3 — це буква «з»?
- Якщо умова виконана, перейти до п. 41, якщо ні — до п. 25.
25. Перевірити: $R_3=R_4$?
- Якщо умова виконана, перейти до п. 26, якщо ні — до п. 36.
26. Виконати: ОМП= N символа — 2? де ОМП — останнє місце переносу.
- Якщо умова виконана, перейти до п. 27, якщо ні — до п. 29.
27. Виконати: $0 \Rightarrow$ ОМП.
 28. Виконати: ЛІЧП — 1 \Rightarrow ЛІЧП, де ЛІЧП — лічильник переносів.
 29. Перевірити: $P+5=0$?
- Якщо умова виконана, перейти до п. 33, якщо ні — до п. 30.
30. Перевірити: R_3 — приголосна?
- Якщо умова виконана, перейти до п. 41, якщо ні — до п. 31.
31. Виконати: 1 \Rightarrow $P+5$.
 32. Перейти до п. 42.
 33. Виконати: N смв. — 1 \Rightarrow МП, де МП — масив переносів.
 34. Виконати: ЛІЧП+1 \Rightarrow ЛІЧП.
 35. Перейти до п. 30.
 36. Перевірити: $P+5=0$?
- Якщо умова виконана, перейти до п. 37, якщо ні — до п. 41.
37. Перевірити: R_3 — це буква «й»?
- Якщо умова виконана, перейти до п. 38, якщо ні — до п. 44.
38. Наступна буква — це буква «о»?
- Якщо умова виконана, перейти до п. 39, якщо ні — до п. 41.

39. Виконати: $N_{\text{СИМВ}} - 1 \Rightarrow \text{МП}$.

40. Виконати: $\text{ЛІЧП} + 1 \Rightarrow \text{ЛІЧП}$.

41. Виконати: $0 \Rightarrow P + 5$.

42. Виконати: $R_1 \Rightarrow R_2$; $R_3 \Rightarrow R_4$.

43. Перейти до п. 13.

44. Перевірити: R_3 — це буква «Ь»?

Якщо умова виконана, перейти до п. 41, якщо ні — до п. 45.

45. Перевірити: $R_1 = R_2$?

Якщо умова виконана, перейти до п. 47, якщо ні — до п. 46.

46. Перевірити: 11 р. $R_1 = 0$?

Якщо умова виконана, перейти до п. 39, якщо ні — до п. 41.

47. Перевірити: 11 р. наступної букви = 0?

Якщо умова виконана, перейти до п. 48, якщо ні — до п. 39.

48. Перевірити: наступна буква — це «Ь»?

Якщо умова виконана, перейти до п. 39, якщо ні — до п. 41.

49. Процес закінчити.

Запропонований алгоритм дає змогу переносити слова з похибкою в 0,3% від загальної кількості переносів. Подібний алгоритм можна застосовувати у задачах, які при розрахунку виключки рядків потребують переносу слова.

ЛІТЕРАТУРА

1. Савинков В. И. Программирование для ЭВМ «Минск-22». М., «Статистика», 1969.

A. S. BERLIN, V. M. SENKIVSKIY, G. Ya. KUSHNIR, V. T. YEFREMENKO

THE MACHINE ALGORITHM OF WORDS' DIVISION FOR THE UKRAINIAN LANGUAGE

Summary

The paper contains the description of the machine rules (algorithms) of words' division for the Ukrainian language. They are a part of rules of the work of general-purpose computer, included in the automatic typesetting programming system. The algorithms are intended for the developing of working programs of «Minsk-22» and «Minsk-32» computer at the preparation of the control program for automatic linecasters.