

2. АНАЛІЗ ЧАСТОТИ ЗУСТРІЧІ ЗНАКІВ У ТЕКСТІ

Кісіль Г.Й., Кризь А.І. - студенти 4 курсу

(Українська академія друкарства)

Науковий керівник: к.т.н., доц. Миклушка І.З.

Існуючі наробики програмного підрахунку частоти зустрічі окремих знаків базуються на таких засадах: перебирання кожного знаку текстового масиву (крок 1); заявлення масиву знаків, які підлягають аналізу (крок 2); при співпаданні знайденого знаку з шуканим знаком спрацьовує лічильник, який збільшує своє значення на одиницю (крок 3). На мові QuickBasic це реалізується таким алгоритмом:

```

L% = LEN (S$)                                     'крок 1
IF L% > 0 THEN
FOR I% = 1 TO L%
SYM$ = MID $(S$, I%, 1)
FOR J% = 32 TO 249                                 'крок 2
IF (J%>63 AND J% ,128) OR (J%>175 AND J%<224) GOTO MITKA
IF SYM$=CHR$(J%) THEN QuanSym%(J%)=QuanSym%(J%)+1 'крок 3
MITKA:
NEXT J%
NEXT I%
END IF

```

Такий аналіз текстового масиву є трудомістким, оскільки вимагає кількості перевірок кожного знаку, що дорівнює величині заявленого масиву (для українських текстів - 155 знаків).

Пропонується інший алгоритм, який полягає в рекурсивному виклику масиву знаків: перебирання кожного знаку текстового масиву (крок 4); виклик з масиву знаків лічильника знайденого знаку, який збільшує своє значення на одиницю (крок 5).

```

L% = LEN(S$)                                     'крок 4
IF L% > 0 THEN
FOR I% = 1 TO L%
SYM$ = MID$(S$, I%, 1) : J% = ASC(SYM$)
QuanSym%(J%) + 1                                 'крок 5
NEXT I%
END IF

```

Робота програми згідно з цим алгоритмом прискорюється в 155 разів, так як відсутня операція перевірки знаку на відповідність знакам із заявленого масиву.