

Güləhməd Məlikov  
Akif Hacıyev  
Azərbaycan Dillər Universiteti

## MÜƏYYƏN MƏTN ƏSASINDA MÜRƏKKƏB ADLARDAN İBARƏT LÜĞƏTİN YARADILMASININ ALQORİTMİ VƏ PROQRAMI

*Açar sözlər:* lüğət, mürəkkəb adlar, alqoritm, proqram

*Keywords:* dictionary, compound names, algorithm, software

*Ключевые слова:* словарь, составные имена, алгоритм, программа

Mətnlərin avtomatik işlənməsi sistemlərində və maşınla tərcümə (MT) problemləri ilə əlaqədar olaraq kompüter lüğətlərinin hazırlanmasının əhəmiyyəti və zəruriliyi müasir dövrümüzün reallığıdır. Formal təhlil sisteminin reallaşdırılması üçün bu cür lüğətlərin olması mühüm şərtidir. Azərbaycan dili üçün tərtib edilmiş kompüter lüğətləri və ya avtomatik lüğətlər (AL) aşağıdakılardan ibarətdir:

- əsaslar AL;
- ən çox işlənən söz-formalar AL;
- birləşmələr AL [1].

“Dilmanc” layihəsi çərçivəsində hazırlanmış Azərbaycan dilindən ingilis və rus dillərinə və əksinə maşın tərcüməsini həyata keçirən sistemlərdə bu tip lüğətlər yaradılmışdır. “Əsaslar avtomatik lüğəti”ndə müəyyən dilin leksik bazasına aid ayrı-ayrı sözlər saxlanıldığı halda, “Söz birləşmələri avtomatik lüğəti”ndə isə müəyyən qaydalara ciddi əməl etməklə iki və daha çox sözdən ibarət olan söz birləşmələri saxlanılır. “Söz birləşmələri avtomatik lüğəti”ndə frazeoloji söz birləşmələri, mürəkkəb adlar, tərkib feillər, ixtisarlar və digər bu kimi həm sabit və həm də sərbəst söz birləşmələri saxlanılır [2].

Elm, texnika və ayrı-ayrı istehsal sahələrinin sürətli inkişafı ilə əlaqədar olaraq “Söz birləşmələri avtomatik lüğəti”nin tez-tez yenilənməyə ehtiyacı olur. Bu korrektə ən çoxu lüğətin “mürəkkəb adlar”la əlaqədar hissəsiylə bağlıdır. Belə ki, onların tərkib hissələri dilin lüğət tərkibinin dəyişməsindən asılı olaraq dəyişikliyə məruz qalır, ümumi model və məna qaldığı halda, birləşməni əmələ gətirən sözlərdən biri başqa bir sözlə əvəz olunur. Məsələn, maarif komissarlığı – maarif nazirliyi, Azərbaycan Dövlət Darülfünü – Azərbaycan Dövlət Universiteti.

Digər tərəfdən unutmamaq olmas ki, mürəkkəb adlar mətnin tərtib olunması gedişində istifadə olunduqca, onların ixtisarlari (abbreviaturaları) hazırlanır və sonrakı proseslərdə mətnin yığcamlığı və sürətli tərtib olunması xatirinə həmin abbreviaturalardan istifadə olunur. Qeyd etmək yerinə düşər ki, abbreviaturaların müvəffəqiyyət qazanmasının digər səbəbi tam yeni bir sözün yaranmasına səbəb olmasıdır, başqa sözlə abbreviaturalar aid olduğu dilin lüğət tərkibini zənginləşdirir. Beləliklə, məndə abbreviaturanın tərtib olunduğu mövqedən sonra mürəkkəb adın əvəzinə müvafiq ixtisardan (abbreviaturadan) istifadə olunduqda MT sistemlərində problem yaranmaması üçün həmin ixtisara uyğun lüğət məqaləsi (abbreviatura, uyğun mürəkkəb ad, çıxış dilində mürəkkəb ad və ona uyğun ixtisar və s.)

“Söz birləşmələri avtomatik lüğəti”nə daxil edilməlidir. Kağızsız texnologiyaların geniş tətbiq olunduğu müasir dövrdə informasiyanın, o cümlədən mətn formasında hazırlanmış məlumatların çox sürətlə artdığını və belə mətnlərdə mürəkkəb adlardan ibarət söz birləşmələrinin mövcud ola biləcəyini nəzərə alsaq, “Söz birləşmələri avtomatik lüğəti”nin vaxtaşırı ənənəvi üsullarla (klaviaturadan əllə verilənlər bazasına daxil etməklə) yeniləşdirilməsinin nə qədər yorucu bir iş olduğunu təsəvvür etmək çətin deyil. Ayrı-ayrı dillər üzrə mətn korpusunun hazırlandığı indiki şəraitdə bu problem özünü daha qabarıq şəkildə büruzə verir [1]. Ona görə də emal olunan mətnə rast gəlinən mürəkkəb adlardan ibarət söz birləşmələrinin AL-ə avtomatik yolla daxil olunması mətnlərin kompüterlə işlənməsi sistemlərində məhsuldarlığı əhəmiyyətli dərəcədə artırma bilər. Bu məqalədə həmin işin alqoritmi və proqramı verilmişdir.

Alqoritmın addımbaaddım təsvirinə keçməzdən əvvəl görülmək üçün mətn formasında izahını verək.

İlk növbədə emal ediləcək mətn formal vasitələrin köməyiylə cümlələr massivi şəklində ifadə olunmalıdır. Sonra hər bir cümlədə mürəkkəb adlardan ibarət söz birləşmələrinin olub-olmadığını müəyyənləşdirmək üçün həmin birləşmənin ixtisarını özündə saxlayan açılan mötərizə – “(” və bağlanan mötərizə – “)” işarələrindən istifadə etmək olar. Əgər cümlə daxilində bu cür konfigurasiyaya rast gəlinərsə, onda mötərizələrlə əhatə olunmuş simvollar ardıcılığının ixtisar (abbreviatura) olub-olmadığını müəyyənləşdirmək üçün onu açılan mötərizə işarəsindən əvvəl gələn sözlərin ilk hərflərinin konkotenasiya olunması (ardıcıl olaraq birləşdirilməsi) nəticəsində alınan formal sözlə (formal sözün hərflərinin sayı mötərizə içərisindəki ixtisarinin hərflərinin sayı ilə təyin olunur) müqayisə etmək lazımdır. Üst-üstə düşmə halında həmin ixtisar və ona aid digər parametrləri özündə saxlayan lüğət məqaləsinə uyğun yazını verilənlər bazasında (VB) yerləşdirmək lazımdır. Əks halda cümlələr massivinin növbəti elementinə keçməklə prosesi başa çatdırmaq lazımdır.

Yaddan çıxarmaq olmaz ki, mürəkkəb adlara uyğun ixtisarlar böyük hərflərlə yazıldığı halda, onu təşkil edən sözlər kiçik hərflərlə də başlaya bilər, məsələn, sistem proqram təminatı (SPT), tətbiqi proqram təminatı (TPT) və s. Bunu nəzərə alaraq açılan mötərizə işarəsindən əvvəl gələn sözlərin ilk hərflərinin konkotenasiya olunması nəticəsində alınan formal sözü müqayisədən əvvəl böyük hərflərə çevirmək lazımdır [3].

Yuxarıdakı izahdan sonra mürəkkəb adlardan ibarət lüğətin yaradılmasının avtomatlaşdırılması üçün emal olunan mətnə formalizasiya elementlərinin olduğunu nəzərə alaraq, kompüterdə realizə alqoritmını aşağıdakı kimi vermək olar:

1. Emal olunacaq mətnin cümlələrindən ibarət birözlü cümlələr massivinin yaradılması (cümlələri bir-birindən ayırmaq üçün “.”, “:”, “?”, “!” və s. işarələrindən istifadə etmək olar);

2. Massivin ilk elementindən başlayaraq cümlə daxilində mötərizə ilə əhatələnmiş ifadənin olub-olmamasının müəyyənləşdirilməsi;

3. Əgər mötərizə ilə əhatələnmiş ifadə yoxdursa, cümlələr massivinin növbəti elementinə, başqa sözlə, 2-ci addıma, əks halda növbəti addıma keçməli;

4. Mötərizə içərisindəki mətn fraqmentinin abbreviatura olub-olmadığının müəyyənləşdirilməsi;

5. Əgər abbreviatura deyilsə, 2-ci addıma, əks halda növbəti addıma keçməli;

6. Cari cümlədəki sözlərdən ibarət sözlər massivi yaratmalı (bu məqsədlə cümlə daxilindəki sözlərin bir-birlərindən ayrılmasını təmin edən boşluq (" ") işarəsindən istifadə etmək lazımdır);

7. Abbreviaturada simvolların sayının müəyyənləşdirilməsi;

8. Sözlər massivindən abbreviaturanı daxilinə alan açılan mötərizə işarəsinə qədər, ardıcıl olaraq ixtisardakı simvolların sayı qədər sözlərin ilk hərflərindən ibarət formal söz yaratmalı;

9. Həmin sözü böyük hərflərə çevirməli və abbreviatura ilə müqayisə etməli;

10. Əgər üst-üstə düşürlərsə, abbreviatura, onun açılışı və digər parametrlərini özündə saxlayan lüğət məqaləsinə uyğun yazını xarici məlumat daşıyıcısında yerləşdirməli, əks halda 2-ci addıma keçməli;

11. Cümlələr massivinin elementləri sona çatana kimi prosesi davam etdirməli.

Bu alqoritmə uyğun Visual Basic dilində tərtib olunmuş proqram aşağıda verilmişdir:

```
Dim cumle, metn, zen, h As String
Dim n, k, l, m, mm, j As Integer
Dim t As Boolean
' Ayırıcı işarələr zənciri
zen = ".,:;!?"
' Mətnin daxil edilməsi
metn = Selection.Text
n = Len(metn)
' Mətnədə cümlələrin sayı
i = 1: j = 0
While i <= n
h = Mid(metn, i, 1)
t = (h = ".") Or (h = ";") Or (h = ":") Or (h = "!") Or (h = "?")
If t = True Then j = j + 1
i = i + 1
Wend
' Ayırıcı işarələrin yerini bildirən və cümlələrdən ibarət massivin təsviri
ReDim prob(j) As Integer, cumlee(j) As String
' Ayırıcı işarələrin mövqeyindən ibarət massivin yaradılması
j = 0
For i = 1 To n
h = Mid(metn, i, 1)
mm = InStr(1, zen, h)
If mm <> 0 Then j = j + 1: prob(j) = i
```

```
Next i
' Cümlələr massivinin yaradılması
cumlee(1) = Left(metn, prob(1) - 1)
For i = 2 To j
cumlee(i) = Mid(metn, prob(i - 1) + 1, prob(i) - prob(i - 1) - 1)
Next i
Open "C:\Users\User\Desktop\lugget" For Output As #10
d = "," 'ayirici simvol
ii = 0
7 ii = ii + 1
If ii > j Then GoTo 9
cumle = cumlee(ii)
n = Len(cumle)
k = 0
For i = 1 To n
If Mid(cumle, i, 1) = " " Then k = k + 1
Next i
ReDim nn(k) As Integer, sozler(k + 1) As String
l = 1: m = 1
While l <= k
nn(l) = InStr(m, cumle, " ")
m = nn(l) + 1
l = l + 1
Wend
sozler(1) = Mid(cumle, 1, nn(1) - 1)
For i = 2 To k
sozler(i) = Mid(cumle, nn(i - 1) + 1, nn(i) - nn(i - 1) - 1)
Next i
sozler(k + 1) = Mid(cumle, nn(k) + 1)
bn = 0: sn = 0
For i = 1 To n
h1 = Mid(cumle, i, 1)
If h1 = "(" Then bn = i 'mötərizənin başladığı mövqe
If h1 = ")" Then sn = i 'mötərizənin qurtardığı mövqe
Next i
If (bn = 0) And (sn = 0) Then GoTo 7
abbr = Mid(cumle, bn + 1, sn - bn - 1)
abbrm = "(" + abbr + ")" 'abbreviaturanı saxlayan mətndəki söz
'abbreviaturanın cümləyə daxil olan sözlər massivinin köməyiylə formalaş-
dırılması
For i = 1 To k + 1
If abbrm = sozler(i) Then m = i: GoTo 2
Next i
2 n1 = Len(abbr)
```

```
For i = 1 To n1
If Asc(Mid(abbr, i, 1)) > 90 Then GoTo 7
Next i
soz = "": fraq = ""
For i = m - n1 To m - 1
soz = soz + Left(sozler(i), 1)
frac = frac + sozler(i) + " "
Next i
If UCase(soz) <> abbr Then GoTo 7
MsgBox (soz)
MsgBox (abbr)
Print #10, abbr, d, frac
GoTo 7
9 Close #10
```

Nümunə olaraq götürülən “Azərbaycan Dillər Universiteti (ADU) Azərbaycanda xarici dil mütəxəssisləri, müxtəlif sahələr və beynəlxalq münasibətlər sahəsində peşəkar mütəxəssislər hazırlayan yeganə Universitetdir. ADU respublikamızın digər ali məktəbləri, o cümlədən dilçi mütəxəssislər hazırlayan Bakı Slavyan Universiteti (BSU) ilə də sıx əməkdaşlıq şəraitində fəaliyyət göstərir. Son vaxtlar ADU xarici ölkələrin müvafiq ali məktəbləri, məsələn, Rusiya Federasiyasının Moskva Dövlət Linqvistik Universiteti (MDLU) ilə də əməkdaşlıq haqqında müqavilə imzalamışdır. Belə ki, MDLU-da Azərbaycan dili və Mədəniyyət Mərkəzi fəaliyyət göstərir” mətn fraqmentinin əsasında tərtib olunmuş proqramın köməyiylə aşağıdakı mürəkkəb adlar lüğətini almış olarıq:

ADU Azərbaycan Dillər Universiteti  
BSU Bakı Slavyan Universiteti  
MDLU Moskva Dövlət Linqvistik Universiteti

### **Ədəbiyyat**

1. Mahmudov M. Kompüter dilçiliyi. Bakı: “Elm və təhsil”, 2013.
2. Fətullayev Ə.B. Azərbaycan-ingilis maşın tərcüməsi sistemi üçün rəqəmsal model-ləşdirmə metodunun işlənilib hazırlanması və tətbiqi // Fiz-riy.f.d. ... dis.avtoref., Bakı, 2006.
3. Məlikov G.Y., Müslümov V.B., Novruzov A.H. Tətbiqi dilçilik məsələlərinin MS Office-də həlli. Bakı: “UniPrint”, 2012.

### **Summary**

#### **The Algorithm and Program of Creating a Complex Dictionary of Compound Names Based on Definite Text**

The Article deals with the dictionary of collocations used in translation from one language to another and the automatic algorithm and program of compound names on the basis of a computer text.

## **Резюме**

### **Создание словаря, содержащего сложные имена на основе конкретного текста алгоритм и программа**

В статье представлен алгоритм, программа для составления части сложных имен часто используемого словаря составных слов автоматически на основе определенного компьютерного текста, который обычно используется в системах перевода на другие языки.

*Rəyçi: dos., fil.f.d. C.Babayev*