Է. Ռ. ՆՂԻԱԶԱՐՅԱՆ

ՄԵՔԵՆԱԿԱՆ ԲԱՌԱՐԱՆՆԵՐԻ ԿԱՌՈՒՑՄԱՆ ՈՐՈՇ ՀԱՐՑԵՐ

Մեքենական բառարան ասելով սովորաբար հասկանում ենք մի սարք, որը կոչված է բառերի վերաբերյալ ինֆորմացիայի պահպանման և տրման համար։ Բառարանը բառահոդվածների մի ամբողջություն է, որը կազմված է՝ ելնելով որոնման կարիքներից։ Բառահոդվածները սովորաբար ունեն հետևյալ տեսքը՝

| A | B |

որտեղ՝ A -ն բառահոդվածի գլխաբառն է, իսկ B -ն՝ գլխաբառին համապատասխան ինֆորմացիան։ Որպես բառահոդվածի գլխաբառ կարող են հանդես դալ բառաձևերը, հիմքերը, արմատները, բառակապակցություններն և նույնիսկ նախադասություններ։ Մեքենական բառարանի անվանումները կախված են այն բանից, թե ինչպիսի լեզվական միավորներ են օդտագործվում որպես գլխաբառեր։

Ելնելով նրանից, թե ինչպիսի տեխնիկական միջոցներով են իրականացվում մեքենական բառարանները, վերջիններս կարելի է բաժանել երեք խմբի.

I Մեքենական բառարաններ, որոնք կառուցված են համընդհանուր հաշվողական մեքենաների վրա ([1], [2])։

II Մեքենական բառարաններ, որոնք կառուցված են ասոցիատիվ հիշողությունների վրա ([3], [4], [5])։

III Մեքենական բառարաններ, որոնք մասնագիտացված ինֆորմացիոն սարքեր են՝ կառուցված մասնագիտական թմբուկների կամ սկավառակների վրա ([6], [7], [8])։

Տեխնիկատնտեսական տեսակետից մեքենական բառարանները բնութագրվում են հիշողության ծավալով, մեկ բառի որոնման միջին արժեքով և ժամանակով, ինչպես նաև բառարանի պարունակության օպերատիվ փոփոխման հնարավորությամբ։

Համընդհանուր հաշվողական մեքենաները ապահովում են մեծ ծավալ ունեցող բառարանների ճկուն կազմակերպում (բառահոդվածի գլխաբառերը կարող են լինել ցանկացած լեզվական միավորներ, բառարանի պարունակության փոփոխության լայն հնարավորություն), սակայն

դրանց արագագործությունը ցածր է, իսկ մեկ բառի որոնման արժեքը՝ բարձր։

Ասոցիատիվ հիշողության վրա կառուցված բառարանները շատ արագագործ են, սակայն փոքր ծավալ ունեն. բացի այդ, որպես գլխաբառեր չեն կարող հանդես գալ ցանկացած լեզվական միավորներ։

Երրորդ տիպի մեքենական բառարանները ունեն բարձր տեխնիկատնտեսական տվյալներ (մեծ ծավալ, մեկ բառի որոնումը ցածր արժեք, բառարանի ճկուն կազմակերպում), սակայն արագագործությունը այնուամենայնիվ ցածր է ([9])։

Հետազոտությունները ցույց են տալիս, որ եթե տեքստում ավելի հաճախ հանդիպող բառերը պահվեն II տիպի մեքենական բառարանում, իսկ մնացածները՝ III-ում, ապա այդպիսի հիբրիդային մեքենական բառարանները կունենան ամենաբարձր տեխնիկատնտեսական տվյալներ [9]։

Այս աշխատանքում ուսումնասիրվում են II տիպի մեքենական բառարանների կազմակերպման հետ կապված խնդիրները այն հաշվով, որպեսզի ապահովվեն ինֆորմացիայի որոնման ավելի բարձր արագագործություն և այդ որոնումը իրականացնող սարքի կառուցվածքի պարզություն։ Այլ կերպ ասած՝ ուսումնասիրվում են II տիպի մեքենական բառարանների օպտիմալ կազմակերպման հարցերը։ Ուսումնասիրվում են նաև III տիպի մեքենական բառարանի օգնությամբ բառերի այբբենական դասավորության հետ կապված հարցերը։

§ 1. Ասոցիատիվ հիշողության վրա կառուցված մեքենական բառարանների կազմակերպման ընդհանուր ճարցեր

Ասոցիատիվ հիշող սարքերի կառուցվածքի և աշխատանքի սկզբունքի հետ կապված հարցերը մանրամասն նկարագրված են [10] աշխատանքում։ Այդ հիշողությունը սովորաբար բաղկացած է երկու մասից՝ ասոցիատիվ և ինֆորմացիոն։ Երբ ասոցիատիվ հիշողությունը օգտագործվում է որպես մեքենական բառարան, ապա ասոցիատիվ մասում գրանցվում են գլխաբառերը, իսկ ինֆորմացիոն մասում՝ դրանց մեկնաբանությունները։ Ասոցիատիվ մասն ունի մատրիցայի տեսք, որն ունի n քանակի տողեր և γ քանակի սյուներ։ Ամեն մի տողում կարող է գրվել միայն մեկ բառ (եթե տառերի քանակը փոքր է γ-ից, ապա դատարկ մնացած սյունները լրացվում են զրոներով)։ Եթե ասոցիատիվ հիշողությունում պահվում են բառաձևեր, ապա մուտքային բառի (որոնվող բառի) մեկնաբանության որոնումը կատարվում է պարզ հարցման միջոցով, այսինքն՝ մուտքային բառը միաժամանակ համեմատվում է հիշողության մեջ պահված բոլոր գլխաբառերի հետ, և որևէ մեկի լիովին

համընկման դեպքում հիշողությունից վերցվում է համապատասխան մեկնաբանությունը։ Հակառակ դեպքում հիշող սարքը ազդանշան է տալիս որոնվող բառի բացակայության մասին։ Եթե հիշողությունում պահվում են բառերի հիմքերը, ապա անհրաժեշտ ինֆորմացիայի որոնումը կատարվում է բարդ հարցման միջոցով, այսինքն հիշողության բազմապատիկ հարցմամբ, մինչև որ գտնվի մուտքային բառի հիմքը։ Ակընհայտ է, որ պահվող հիմքերը և կատարվող հարցումները պետք է բխեն ընտրված ձևաբանական վերլուծության եղանակից։

Մեր առջև դրված խնդրի լուծման համար նախընտրել ենք [11] աշխատանքում նկարագրված ձևաբանական վերլուծությունը, ըստ որի մուտքային բառի համար հիմքերի բառարանից անհրաժեշտ է ընտրել այն ամենաերկար հիմքը, որը սկսած բառի սկզբից, լրիվ ընդգրկվում է մուտքային բառի մեջ։ Այս ալգորիթմի իրականացման համար ասոցիատիվ հիշողության վրա կառուցված մեքենական բառարանը պետք է հնարավորություն ունենա՝

ա) անտեսել կամ ջնջել մուտքային բառի որևէ տառը կամ տառախումբը,

բ) տարբերակել հետևյալ դեպքերը՝

— չկա ոչ մի ընդգրկվող հիմք ($\nu_0 = 1$),
— կա միայն մեկ ընդգրկվող հիմք ($\nu_1 = 1$),
— կան մեկից ավելի ընդգրկվող հիմքեր ($\nu_2 = 1$):

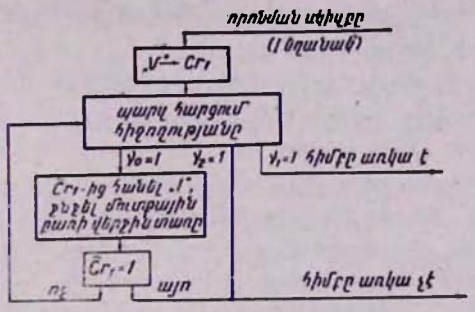
Նշված ալգորիթմով մուտքային բառի հիմքի որոնումը ասոցիատիվ հիշողությունում կարելի է իրականացնել մի քանի եղանակներով։ Բերենք այդ եղանակներից ամենահիմնականները։

§ 1.1. Հիմքի որոնումը մուտքային բառի վերջին տառի ջնջման միջոցով (եղանակ I)։

Հիմքի որոնումը սկսվում է նրանից, որ մուտքային բառը, որի տառերի քիվը ընդունվում է որպես γ-ի հավասար, համեմատվում է հիշողության մեջ պահված բոլոր հիմքերի հետ։ Եթե համեմատությունից պարզվում է, որ ա) $\nu_0 = 1$, ապա ջնջվում է մուտքային բառի վերջին տառը (այն կատարվում է ըստ Cr_3 հաշվիչի պարունակության) և համեմատումը կրկնվում է։ Այդ շարունակվում է այնքան անգամ, մինչև որ ստացվեն $\nu_1 = 1$ կամ $\nu_2 = 1$ ազդանշանները, կամ էլ պարզվի, որ մուտքային բառից մնացել է միայն մեկ տառ (ընդունվում է, որ հիմքը չի կարող ունենալ միայն մեկ տառ)։ Առաջին դեպքը ($\nu_1 = 1$) նշանակում է, որ հիմքը առկա է, իսկ մյուսները՝ հիմքը առկա չէ։

Նկ. 1-ում պատկերված է I եղանակով կատարվող հիմքի որոնման ալգորիթմի ընդհանուր բլոկ-սխեման։ Այս աշխատանքում նպատակա-

հարմար չենք համարում ներկայացնել որոնումը իրականացնող ավտո-



Նկ. 1

մատ սարքի դրաֆր, սակայն անհրաժեշտ է նշել, որ այդ ավտոմատի ներքին վիճակների քանակը հավասար է 2-ի։

Այստեղ մեկ հիմքի որոնման համար անհրաժեշտ հարցումների միջին քանակը կարելի է որոշել հետևյալ բանաձևով՝

$$n_{01}=1\cdot f_v+2f_{v-1}+\cdots+(v-1+1)f1+\cdots+(v-1)f_2 \qquad (1$$

$$n_{01}=\sum_{i=2}^{v}(v-1+1)f1 \quad (1'),$$

որտեղ f_1 — տեքստում այն մուտքային բառերի հանդես դալու հավանականությունն է, որոնց հիմքերը բաղկացած են i թվով տառերից։

§ 1.2. Հիմքի որոնումը մուտքային բառի սկղբի տառերի համեմատման միջոցով (եղանակ II)։

Հիմքի որոնումը այս եղանակով սկսվում է մուտքային բառի և բառարանում պահվող հիմքերի առաջին երկու տառերի համեմատմամբ։ Դրա համար անհրաժեշտ է անտեսել ինչպես մուտքային բառի, այնպես էլ բառարանում դրված հիմքերի մնացած տառերի առկայությունը։

Եթե առաջին երկու տառերի համեմատությունից ստացվում է, որ

1) $v_0=1$, ապա որոնելի հիմքը առկա չէ,

2) $v_2=1$, ապա համեմատվող տառերի թիվը մեկով ավելացվում է և համեմատությունը կրկնվում է մինչև $v_0=1$ կամ $v_1=1$ ազդանշանների ստացումը։

Եթե ա) $v_1=1$, ապա հնարավոր է, որ ընդգրկվողը լինի որոնելի հիմքը։ Այդ հաստատելու համար անհրաժեշտ է հիմքը ամբողջու-

---

* Դրա համար պետք է կատարել Π_1 գործողությունների հաջորդականությունը՝

ա) կարդալ ընտրված հիմքի ինֆորմացիայի այն մասը, որտեղ գրված է հիմքերի տառերի քիվը և դրել Cr_2 հաշվիչում,

բ) Cr_1 և Cr_2 հաշվիչների պարունակություններից որոշել, թե քանի տառ պետք է համեմատել։

թյամբ համեմատել մուտքային բառի հետ՝, եթե համեմատությունից ստացվում է $v_1 = 1$, ապա ընտրված հիմքը որոնելին է, իսկ $v_0 = 1$-ը նշանակում է, որ հիմքը առկա չէ։

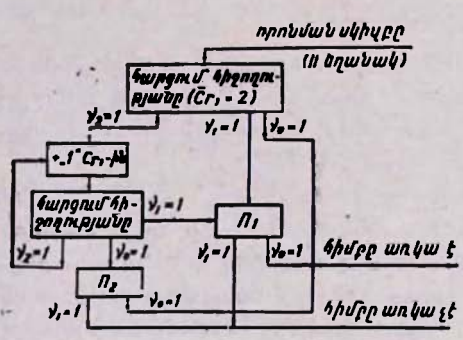
Եթե բ) $v_0 = 1$, ապա այն դեռևս չի նշանակում, թե հիմքը առկա չէ, քանի որ չհամընկնելը կարող է հետևանք լինել այն բանի, որ հիմքերը արհեստականորեն երկարացված են (վերջում զրոներ գրելու ճանապարհով)։ Ուստի անհրաժեշտ է ջնջել մուտքային բառի վերջին տառերը և կատարել պարզ հարցում հիշողությանը Π_0։ Այս դեպքում $v_0 = 1$-ը կնշանակի հիմքի բացակայություն, իսկ $v_1 = 1$-ը՝ առկայություն։

Այս եղանակի դեպքում մեկ հիմքի որոնման համար անհրաժեշտ հարցումների միջին քանակը որոշվում է հետևյալ բանաձևով՝

$$n_{o2} = 2F_2 + 3F_3 + \cdots + iF_i + \cdots + vF_v \qquad (2)$$

$$n_{o3} = \sum_{i=2}^{v} iF_i \qquad (2^1),$$

որտեղ F_i-ն մուտքային բառի սկզբնական տառերի հերթական համեմատության ժամանակ $v_i = 1$ ազդանշան ստանալու հավանականությունն է։



Նկ. 2

Երկրորդ եղանակով կատարվող հիմքի որոնման ալգորիթմի ընդհանուր բլոկ-սխեման պատկերված է նկ. 2-ում։ Այս եղանակով հիմքի որոնումը իրագործելու համար անհրաժեշտ է հինգ ներքին վիճակներով ավտոմատ, երկու հաշվիչներ (Cr_1, Cr_2), ինչպես նաև անտեսումն ապահովող սխեմաներ։

§ 1.3. Հիմքի որոնման I և II եղանակների համեմատական վերլուծությունը։

Համեմատելով հիմքի որոնման վերոհիշյալ երկու եղանակները ըստ օգտագործվող սարքերի, կարելի է ասել, որ II եղանակի օգտագործումը շահավետ չէ, սակայն վերջնական եզրակացության գալու համար ան-

հրաժեշտ է համեմատել դրանց արագագործությունը։ Դրա համար (1) և (2) արտահայտություններից անհրաժեշտ է հաշվել n_{o_1}-ը և n_{o_2}-ը։ Նման հաշվումներ կատարելու համար անհրաժեշտ է ունենալ հիմքերի հաճախականությունների բառարան։ Վերջինս (շուրջ 900 հիմքի համար) ստացել ենք մաթեմատիկական տերմինների հաճախականության բառարանից [12]։

Ստորև շարադրում ենք հիմքերի հաճախականությունների բառարանի կազմման և f_i-ի ու F_i-ի որոշման մեթոդը։

Ընդունենք, որ հաճախականության բառարանը ունի հետևյալ տեսքը՝

№	Բառեր	Հանդես գալու հավանակ.
1	по	P_1
2	представлять	P_2
3	представляться	P_3
4	преимуще тв)	P_4
5	преимущественно	P_5
6	при	P_6
7	применить	P_7
8	применять	P_8

Փոխարինելով բառերը իրենց հիմքերով՝ կստանանք հիմքերի հաճախականության բառարան

№	Բառեր	Հանդես գալու հավանակ.
1	по	P_1
2	представ	$P_2 + P_3$
3	преимуществ	$P_4 + P_5$
4	при	P_6
5	примен	$P_7 + P_8$

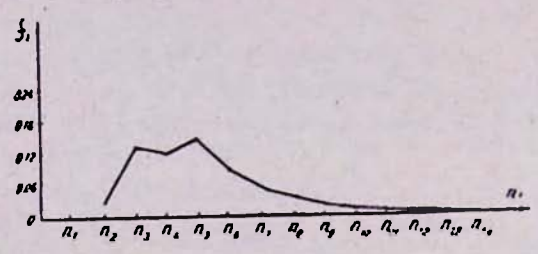
Ստացված աղյուսակի հիման վրա, թեքվող ձևերի համար հաշվենք f_i-ի և F_i-ի արժեքները։ Կստանանք՝ $f_{11} = P_4 + P_5$; $f_{10} = f_9 = 0$; $f_8 = P_2 + P_3$; $f_7 = 0$; $f_6 = P_7 + P_8$; $f_5 = f_4 = f_3 = f_2 = 0$; $F_2 = F_3 = 0$; $F_4 = P_2 + P_2 + P_4 + P_5 + P_6 + P_7 + P_8$,

Մաթեմատիկական տերմինների բառարանի հիման վրա ստացված i-ի և F_i-ի արժեքների բաշխվածությունների գրաֆիկները համապատասխանաբար բերված են 3-րդ և 4-րդ նկարներում։

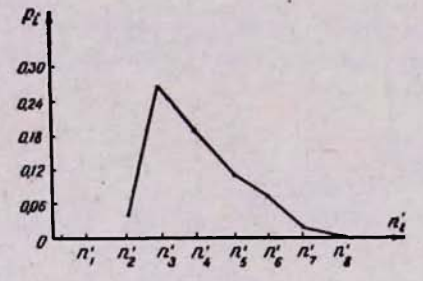
Տեղադրելով f_i-ի և F_i-ի արժեքները (1) և (2) բանաձևերում, կստանանք՝

$n_{o1} = 6,7$ և $n_{o2} = 2,7$

Այսպիսով՝ ստացվեց, որ որոնման II եղանակը ավելի արագագործ է։ Այն բացատրվում է նրանով, որ F_i-ն բաշխված է համեմատաբար ավելի նեղ տիրույթում, քան f_i-ն: Այստեղ հարց է ծագում, ինչպես էլ ավելի բարձրացնել հիմքի որոնման արագագործությունը։



Նկ. 3



Նկ. 4

Մեր հաշվումները ցույց են տալիս, որ եթե հիշողությանը առաջին հարցումը կատարվի ոչ թե F_i-ի բաշխվածության տիրույթի սկզբից, այլ նրա «կենտրոնից», ապա հիմքի որոնման արագությունը զգալիորեն կաճի։ Այս նոր եղանակը նկարագրված է հաջորդ պարագրաֆում։

§ 1.4. *Հիմքի որոնումը «կենտրոնից» (եղանակ III)։*

Այս եղանակով հիմքի որոնումը սկսվում է մուտքային բառի սկզբի $i (i>2)$ քանակի տառերի համեմատմամբ՝ հիմքերի համապատասխան տառերի հետ։ Եթե ստացվում է՝

(1) $v_0 = 1$, ապա մուտքային բառի մնացած տառերը չնչվում են, և հետագա որոնումը կատարվում է ըստ I եղանակի.

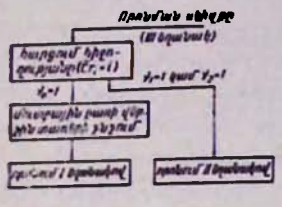
2) $v_2 = 1$ կամ $v_1 = 1$, ապա որոնումը կատարվում է ըստ II եղանակի։

5-րդ նկարում բերված է ալգորիթմի բլոկ-սխեման։

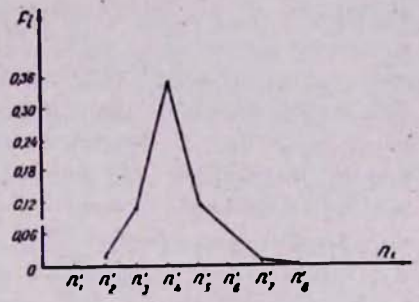
III եղանակի դեպքում մեկ հիմքի որոնման համար անհրաժեշտ հարցումների միջին քանակը որոշվում է հետևյալ բանաձևով՝

$$n_{o3} = 2F^1 + 3F_1 + \cdots + \\ + 2F_{l-1} + 3F_{l-2} + \cdots \qquad (3)$$

Եթե i-ին տանք հետևյալ արժեքները՝ 3,4,5 և այլն, ապա կտեսնենք, որ n_{03}-ը նախ նվազում է, ապա՝ աճում, n_{03}-ը իր ամենափոքր արժեքը՝ 1,8-ը ընդունում է, երբ i=4-ի։ Այս դեպքի համար F_1-ի բաշխվածության գրաֆիկը բերված է 6-րդ նկարում։



Նկ. 5



Նկ. 6

Այսպիսով պարզեցինք, որ III եղանակը ամենաարագագործն է, քանի որ նա գերազանցում է I եղանակի արագագործությանը 3,7 անգամ, իսկ II-ի՝ 1,5 անգամ։

Աղյուսակ 1-ում բերված են հիմքի որոնման քննարկված 3 եղանակները բնութագրող հիմնական տվյալները։ Համեմատելով այդ տվյալները և հենվելով այն բանի վրա, որ՝

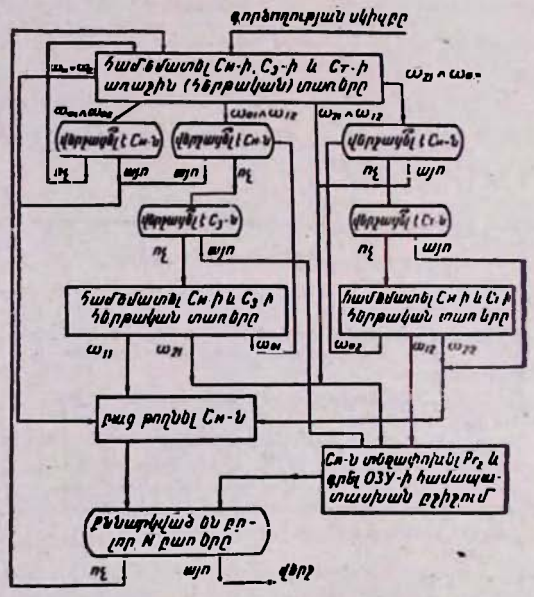
Աղյուսակ №

հիմքի որոնման եղանակը	Օգտագործվող հիմնական սարքերը	Արագագործությունը
I եղանակ	Ներքին 2 վիճակով մեկ ավտոմատ, մեկ հաշվիչ (Cr_1), մուտքային բառի տառերը չընջող սարք	$n_{01}=6,7$
II եղանակ	Ներքին 3 վիճակով մեկ ավտոմատ, երկու հաշվիչ (Cr_1, Cr_2), մուտքային բառի տառերի գոյությունը անտեսող սարք	$n_{02}=2.7$
III եղանակ	Ներքին 6 վիճակով մեկ ավտոմատ, երկու հաշվիչ (Cr_1, Cr_2), մուտքային բառի տառերի գոյությունը անտեսող սարք, մուտքային բառի տառերը չնջող սարք	$n_{03}=1,8$

ա) որոնման համար օգտագործվող սարքավորումը Cr_1, Cr_2, մուտքային բառերի գոյությունը անտեսող սարքը և այլն օգտագործվում են լեզվավիճակագրական որոշ աշխատանքների մեքենայացման համար (13),

բ) III եղանակի կիրառման դեպքում օգտագործվող լրացուցիչ սարքավորումը ընդհանուր օգտագործվող սարքավորման չնչին մասն է կազմում։ Մեր կարծիքով նպատակահարմար է օգտագործել III եղանակը, քանի որ այն ամենաարագագործն է։

Գոյություն ունեն բավականին քվով ալգորիթմներ, որոնց օգնությամբ բառերը հաշվողական մեքենայի միջոցով դասավորվում են ըստ այբուբենի [14], [15]։ Այդ ալգորիթմների ուսումնասիրությունը ցույց է տալիս, որ 3-րդ տիպի մեքենական բառարանի միջոցով բառերը ըստ այբուբենի դասավորելու համար նպատակահարմար է օգտագործել այն ալգորիթմը, որն իրականացվում է «տրված բառին ամենամոտ մեծ բառի որոնումը» գործողության ցիկլիկ կատարման միջոցով։ Տրված բառերի բազմության մեջ ցանկացած C_1 բառին ամենամոտ մեծ բառ ասելով հասկանում ենք այդ բազմության այն C_2 բառը, որը բազմության այբբենական դասավորման դեպքում կդրվեր C_1-ից անմիջապես հետո։ Այս գործողության կատարման ալգորիթմի ընդհանուր բլոկ-սխեման ներկայացված է նկ. 7-ում։

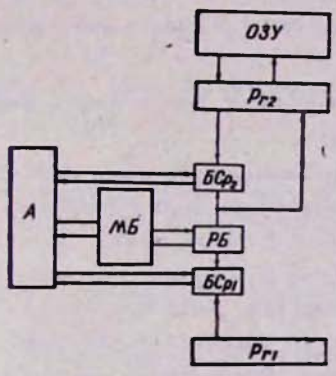


Նկ. 7

Ենթադրենք, որ չկարգավորված բառերի բազմությունը գրված է մագնիսական թմբուկի (МБ)m շավիղների վրա։ Ընդ որում բառերը գրված են անմիջապես մեկը մյուսի ետևից և միմյանցից բաժանվում են «Z» տառով, իսկ բազմության վերջը որոշվում է „Z_1“ տառով։ 3-րդ տիպի մեքենական բառարանի այն մասի ընդհանուր բլոկ-սխեման,

որի միջոցով իրականացվում է չկարգավորված բառերի դասավորումը ըստ այբուբենի, պատկերված է նկ. 8-ում։

Այս գործողությունը հիմնականում կատարվում է երեք ռեգիստրների (ՐԲ, Pr_1, Pr_2), երկու համեմատող բլոկների ($БСр_1$, $БСр_2$), օպերատիվ հիշողության (ՕԶՎ) և ղեկավարող սարքի (Ա) օգնությամբ։ Պայմանավորվենք Pr_1-ում գտնվող բառն անվանել տրված բառ ($C_з$), Pr_2-ում գտնվողը՝ ընթացիկ բառ (Cr), իսկ մագնիսական թմբուկից կարդացվող և վերոհիշյալ բառերից որևէ մեկի հետ համեմատվող բառը՝ հետազոտվող բառ (Cu)։



Նկ. 8

Գործողության սկզբում որոնվում է չկարգավորված բառերի բազմության ամենափոքր բառը։ Դրա համար որպես տրված բառ վերցվում է գոյություն չունեցող ամենափոքր բառը, այսինքն Pr_1-ը թողնվում է դատարկ, իսկ որպես ընթացիկ բառ վերցվում է գոյություն չունեցող ամենամեծ բառը։ Գործողությունը սկսվում է չկարգավորված բառերից ամենաառաջին բառի առաջին տառի համեմատությամբ Pr_1-ի և Pr_2-ի բառերի առաջին տառերի հետ։ Համեմատումը Pr_1-ի հետ կատարվում է $БСр_1$-ի, իսկ Pr_2-ի հետ՝ $БСр_2$-ի միջոցով։ Ընդունենք, որ $БСр_1$-ը համեմատության ընթացքում տալիս է հետևյալ ազդանշանները՝

$\omega_{01} = 1$, եթե համեմատվող տառերը նույնն են,

$\omega_{\cdot\cdot} = 1$, եթե հետազոտվող բառի տառը փոքր է տրված բառի համապատասխան տառից,

$\omega_{21} = 1$, եթե մեծ է։

Ընդունենք նաև այն, որ $БСр_2$-ը տալիս է համանման ազդանշաններ՝ $\omega_{02} = 1$. եթե համեմատվող տառերը նույնն են,

$\omega_{12} = 1$, եթե հետազոտվող բառի տառը փոքր է ընթացիկ բառի համապատասխան տառից,

$\omega_{22}=1$, եթե մեծ է։

Պարզ է, որ չկարգավորված բառերի բազմության հենց առաջին բառի առաջին տառի համեմատումից կստացվեն ω_{21} և ω_{12} ազդանշանները։ Սա նշանակում է, որ հետազոտվող բառը մեծ է տրված բառից և միևնույն ժամանակ փոքր է ընթացիկ բառից, այսինքն որոնելի բառ է, ուստի նա փոխարինում է ընթացիկ բառին և միևնույն ժամանակ գրվում է օպերատիվ հիշողության (k+1) բջջում։

Հաջորդ բառը (բառերը) համանման ձևով համեմատվում է տրված և ընթացիկ բառերի հետ։ Եթե հետազոտվող բառի առաջին տառի համեմատությունից ստացվում են հետևյալ ազդանշանները՝

1. ω_{21} և ω_{12}, ապա հետազոտվող բառը փոխարինում է ընթացիկ բառից,

2. ω_{22}, ապա հետազոտվող բառը բաց է թողնվում, քանի որ նա մեծ է ընթացիկ բառից,

3. ω_{21} և ω_{02}, ապա հետազոտվող բառը մեծ է տրված բառից, սակայն անհրաժեշտ է պարզել նրա հարաբերությունը ընթացիկ բառի հետ։

Դրա համար վերցվում է հետազոտվող բառի հաջորդ տառը։ Եթե նա «Z»-է (որը նշանակում է, թե բառը վերջացել է), հետազոտվող բառը փոխարինում է ընթացիկին, քանի որ նրանից փոքր է։ Հակառակ դեպքում (երբ այն «Z» չէ) վերցվում է ընթացիկ բառի հաջորդ տառը։ Եթե վերջինս «Z» է, ապա հետազոտվող բառը մեծ է ընթացիկ բառից, ուստի վերջինս բաց է թողնվում։ Հակառակ դեպքում նրանք համեմատվում են։ Եթե համեմատումից ստացվում է՝

ա) 02, ապա 3-րդ կետում գրանցված բոլոր գործողությունները կրկնվում են,

բ) ω_{12}, ապա ընթացիկ բառը փոխարինվում է հետազոտվող բառով,

գ) ω_{22}, ապա հետազոտվող բառը բաց է թողնվում։

Համեմատելով մնացած բառերը, Pr_2-ում կստացվի գոյություն չունեցող ամենափոքր բառին ամենամոտ մեծ բառը, տվյալ դեպքում՝ չկարգավորված N թվով բառերից ամենափոքրը։

Այժմ ուսումնասիրենք ամենաընդհանուր դեպքը, երբ որպես տրված բառ հանդես է գալիս կարգավորված բառերի բազմության ցանկացած i-րդ բառը ($1 \leqslant i \leqslant N$)։ Սա նշանակում է, որ այդ i-րդ բառը գտնվում է Pr_1-ում և գրված է օպերատիվ հիշողության (k+i)-րդ բջջում, իսկ Pr_2-ում գտնվում է i-րդ բառից մեծ ցանկացած մեծ բառ։

Գործողությունը սկսվում է չկարգավորված բառերի բազմության ամենաառաջին բառի առաջին տառի համեմատությամբ, ինչպես տըր-

ված, այնպես էլ ընթացիկ բառերի համապատասխան տառերի հետ։ Եթե համեմատությունից ստացվում է՝

4. ω_{11} կամ ω_{22}, ապա հետազոտվող բառը բաց է թողնվում, քանի որ այն փոքր է տրված բառից կամ էլ մեծ է ընթացիկ բառից,

5. ω_{01}, և ω_{121}, ապա հետազոտվող բառը փոքր է ընթացիկ բառից, սակայն նրա հարաբերությունը տրված բառի նկատմամբ անորոշ է։

Այդ անորոշությունը պարզելու համար կարդացվում է հետազոտվող բառի հաջորդ տառը։ Եթե նա «Z» է, ապա հետազոտվող բառը փոքր է (կամ էլ հավասար է), ուստի բաց է թողնվում։ Հակառակ դեպքում համանման քննության է ենթարկվում տրված բառի համապատասխան տառը. դրական պատասխանի դեպքում հետազոտվող բառը փոխարինում է ընթացիկ բառին և դրվում (k+i+1) բջջում։ Հակառակ դեպքում այդ տառերը համեմատվում են։ Եթե համեմատումից ստացվում է՝

դ) ω_{01}, ապա 5-րդ կետում գրանցված բոլոր գործողությունները կրկնվում են,

ե) ω_{12}, ապա հետազոտվող բառը բաց է թողնվում,

զ) ω_{21}, ապա հետազոտվող բառը փոխարինում է ընթացիկ բառին,

6. ω_{01} և ω_{02}, ապա կարդացվում է հետազոտվող բառի հաջորդ տառը։

Եթե այն «Z» է, ապա հետազոտվող բառը բաց է թողնվում, քանի որ նա փոքր է (հնարավոր է և հավասար է) տրված բառից։ Հակառակ դեպքում այն համեմատվում է ինչպես տրված, այնպես էլ ընթացիկ բառի համապատասխան տառի հետ։ Եթե նորից ստացվի ω_{01} և ω_{02}, ապա 6-րդ կետը կրկնվում է։ Այս պրոցեսը կշարունակվի այնքան ժամանակ, մինչև պարզվի հետազոտվող տառի «Z» լինելը, կամ էլ ω_{01} և ω_{02} ազդանշանից տարբեր ազդանշան (որոնց մասին խոսել ենք վերևում) ստանալը։

Այսպիսով, չկարգավորված բոլոր բառերը դիտարկելուց հետո Pr_0-ում կստանանք i-րդ բառին ամենամոտ մեծ բառը, որը և գրվում է (k+i+1) բջջում։

Կատարելով այս դասդասման գործողությունը N անգամ, օպերատիվ հիշողության (K+1) բջջից մինչև (K+H) բջիջը, այբբենական կարգով կգրանցվեն հետազոտվող բոլոր N բառերը։

ԳՐԱԿԱՆՈՒԹՅՈՒՆ

1. *Медведев А. А.* Поиск по словарю с глобальным размещением словарных единиц. Труды III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации, т. 2, М., 1967.

2. *Betz B., Hoffman W.*, The use of a random access device for dictionary lo okup — In: Research in machine translation. — Russian → English. Wayne sta te univ./ Fifth annual report). Detroit, 1963.
3. *Гутенмахер Л. И.*, Автоматический словарь, Авт. свид. № 146375, 24.06. 1961. Бюлл. изобр., 1962, № 10.
4. *Стецюра Г. Г.*, Автоматическое поисковое устройство с обратной связью. Доклад на конференции по машинному переводу и автоматическому чтению текста. ВИНИТИ АН СССР, вып. 4, 1961.
5. *Кулаков Л. В.*, Словарное запоминающее устройство для ЭЦВМ, НТИ, №6, 1965.
6. *Оттингер А. Г.*, Проект автоматического русско-английского технического словаря. В сб. «Машинный перевод», М., 1959.
7. *Вахабов В. К., Шереметьев И. К.* и др., Специализированная цифровая машина для информационного поиска. Доклады на III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации, ВИНИТИ, т. III, М., 1967.
8. *Вартанян Н. В., Егиазарян Э. В., Урутян Р. Л.*, Организация словарей машины «Гарни». Математические вопросы кибернетики и вычислительной техники. Труды ВЦ АН Арм.ССР и ЕрГУ, вып. 7, 1972.
9. *Егиазарян Э. В.*, Некоторые вопросы построения автоматических словарей для машинного перевода. В сб. «Структурный анализ текста», Институт языка им. Р. Ачаряна, 1979.
10. *Крайзмер Л. П., Бородаев Д. А.* и др. Ассоциативное запоминающее устройство, Л., 1967.
11. *Кулагина О. С.*, О машинном переводе с французского языка на русский. I словарь. Проблемы кибернетики, вып. 3, 1960.
12. *Тер-Мисакянц З. Т.*, Частотный словарь математической лексики, Ереван, 1973.
13. *Егиазарян Э. В.*, Применение «гибридного» автоматического словаря для автоматизации некоторых лингвостатистических работ. В сб. «Вопросы анализа текста». Институт языка им. Р. Ачаряна АН Арм. ССР, 1975.
14. *Лавров С. С., Гончарова Л. И.* Автоматическая обработка данных. Хранение информации в памяти ЭВМ, М., 1971.
15. *Селетков С. Н., Волков Б. Т.* Хранение и поиск данных в ЭВМ, М., 1971.