

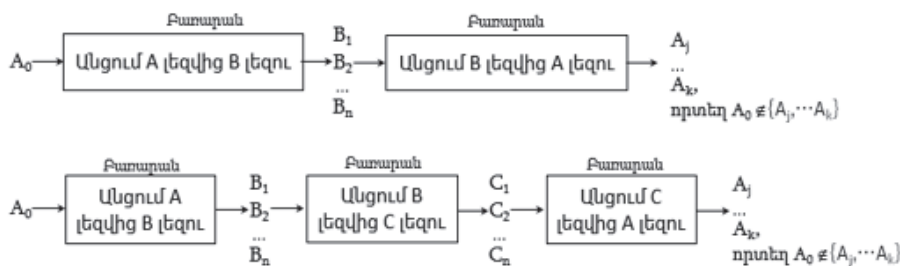
Օ.Վ. ՄԱՆՈՒԿՅԱՆ, Ա.Ս. ՄԱՆՈՒԿՅԱՆ, Է.Ն. ՄԱՆՈՒԿՅԱՆ

ԲԱՑԱՏՐԱԿԱՆ ԲԱՌԱՐԱՆՆԵՐԻ ՄՇԱԿՄԱՆ ISMA ՏԵԽՆՈԼՈԳԻԱ

Առաջարկվում է ցանցային գիտելիքների շտեմարան՝ որպես բացատրական-թարգմանական բառարանների կազմակերպման ժամանակակից միջոց: Նկարագրվում է բազմալեզու թարգմանական-բացատրական բառարանների կազմակերպման ISMA տեխնոլոգիան, և առաջարկվում էքսպերտային համակարգի կազմակերպման սխեմա, որի նպատակն է օժանդակել ցանցային գիտելիքների շտեմարանի սկզբունքով աշխատող բացատրական բառարանների ստեղծմանը:

Առանցքային բառեր. գիտելիքների շտեմարան, բառարանային հոդված, բացատրական բառարան, էքսպերտային համակարգ, տվյալների ինվերս ներկայացում:

1. Բառարանագիտության որոշ խնդիրներ, որոնք հանդես են գալիս թվային միջավայրում: Վերջին տարիներին համացանցում հայտնվեցին հայերեն բազմաթիվ բառարաններ, որոնք հիմնականում թղթային բառարանների պատճեններն են: Համացանցում դրանց համադրումն ի հայտ է բերում բառարաններում առկա անորոշություններ: Օրինակ (նկ. 1), երբ անցնում ենք A լեզվից B լեզու (օգտվելով $A \rightarrow B$ անցման բառարանից), որևէ A_0 ընթացիկ բառի համար ստանում ենք B_1, \dots, B_n թարգմանություններ: Սակայն երբ $B \rightarrow A$ բառարանով փորձում ենք ստանալ այդ ցանկում բերված B_i բառի թարգմանությունը հետ դեպի A լեզու, ապա ստացված A_1, \dots, A_n բառացանկում հաճախ տեղ չի գտնում սկզբնական A_0 բառը: Նման անհամաձայնություններն ավելի ակնհայտ են դառնում, երբ օգտագործում ենք 3 բառարաններ՝ $A \rightarrow B, B \rightarrow C, C \rightarrow A$:



Նկ. 1

Նշված խնդիրը լուծելու, բառարանագրության մեջ նկատվող թերությունները չկրկնելու համար այս աշխատանքում առաջարկվում են բառարանների ստեղծման նոր տեխնոլոգիա, որը թույլ կտա նվազագույնի հասցնել նման

անորոշությունները: Առաջարկվող միասնական բառարանը հարմար է նաև տարբեր ոլորտների բառերը մեկ ընդհանուր բառարանում ինտեգրելու համար:

2. Թվային միջավայրին հարմարեցված բառարանակազմության նոր տեխնոլոգիայի մասին: Այս աշխատանքում ցանկացած բառարան առաջարկվում է դիտել որպես տվյալների շտեմարան: Ուղիղ պահման ժամանակ յուրաքանչյուր ինֆորմացիոն օբյեկտ (ԻՕ) շտեմարանում ներկայացվում է ինֆորմացիայի կրիչի մեկ հատվածով, իսկ տվյալ օբյեկտի հատկանիշը՝ այդ հատվածի որոշակի դաշտով:

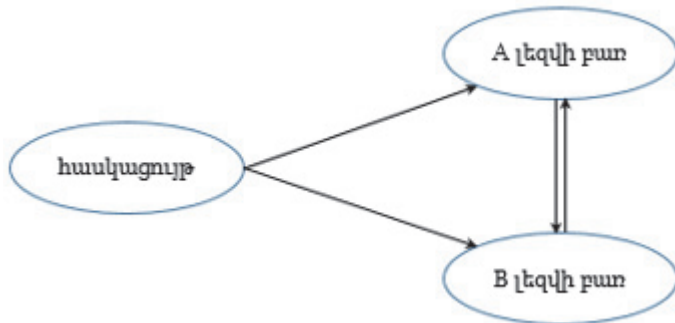
Ինվերս պահման ժամանակ յուրաքանչյուր հատկանիշ ներկայացվում է մեկ գրառումով, որի մեջ պահվում է այդ հատկանիշն ունեցող ինֆորմացիոն օբյեկտների համարների ցուցակը:

Թվային բառարանների կազմակերպման դեպքում ԻՕ-ն թարգմանվող լեզվի բառն է, իսկ հատկանիշը՝ դրա որևէ իմաստն արտահայտող բացատրությունը: Թարգմանական բառարանների թվայնացման դեպքում այդ իմաստին կցվում է նույն իմաստն արտահայտող բառը թարգմանվող լեզվով: Ներկայումս թվայնացված բառարանները կարելի է դիտել որպես ուղիղ ձևով կազմակերպված տվյալների շտեմարան:

Աշխատանքում առաջարկվում է բառարանը կազմակերպել որպես ինվերս ձևով պահվող տվյալների շտեմարան, այսինքն նրա կազմակերպական միավորն արդեն բառի իմաստն է՝ առարկայական աշխարհի որևէ հասկացույթը (այսուհետ՝ կոնցեպտ), իսկ տվյալ կոնցեպտին կցվում են թարգմանվող լեզվի այն բոլոր բառերը, որոնք ներկայացնում են համապատասխան հասկացույթի տվյալ իմաստը: Քանի որ առարկայական աշխարհը համարվում է ընդհանուր թե՛ թարգմանվող և թե՛ թարգմանվող լեզուների համար, այսինքն մի լեզվով սահմանված կոնցեպտները պետք է առկա լինեն երկրորդ լեզվով սահմանված կոնցեպտների բազմության մեջ, ապա շտեմարանի ցանկացած կոնցեպտին կարելի է կցել նաև դրա իմաստը երկրորդ լեզվով արտահայտվող բառերի՝ հոմանիշների ցուցակում:

Այժմ դիտարկենք նման բառարանի ձևավորման գործընթացը:

Քանի որ այստեղ որոշիչը կոնցեպտն է, ապա սկզբից այն սահմանվում է՝ ելնելով առարկայական աշխարհում իր գրաված դիրքից և հարաբերություններից: Հաջորդ քայլում կոնցեպտին կցվում են նրա իմաստն արտահայտող բառերն A լեզվով և հետո՝ նոր նրա իմաստն արտահայտող բառերը B լեզվով: Այսինքն A լեզվի բառերը B լեզվով թարգմանելիս ակտիվացվում է նրանց միջնորդ կոնցեպտը (նկ. 2):



Նկ. 2

Ներկայումս օգտագործվող բառարաններում թե՛ բառարանի ձևավորման և թե՛ նրա շահագործման գործընթացում վերոնշյալ կոնցեպտը հստակ սահմանված չի լինում, հենց դա էլ առիթ է տալիս երկլեզու բառարաններում սխալների առաջացման: Փորձը ցույց է տալիս, որ կոնցեպտի առկայությունը թե՛ շտեմարանի ձևավորման և թե՛ բառային թարգմանության գործընթացում նպաստում է $A \rightarrow B$ անցման սխալների նվազեցմանը: Այսպիսով, բառարանի ստեղծումը սկսվում է կոնցեպտների սահմանումով և նրանց շտեմարանի ստեղծումով: Կոնցեպտներն իրենց կապերով-հարաբերություններով կարելի է դիտել որպես առարկայական աշխարհի կոնցեպտուալ մոդել, որին ինֆորմատիկայում անվանում են գիտելիքների ցանցային շտեմարան:

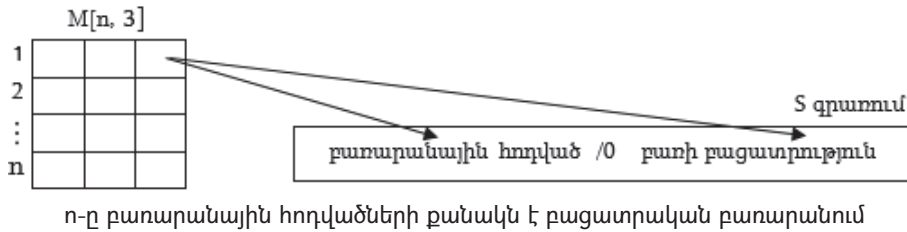
3. Թվային միջավայրում նոր տեխնոլոգիայով կազմված բառարաններում բառերի, նրանց իմաստների ներկայացման մի քանի մեթոդներ: Ելնելով վերը շարադրվածից, առաջարկվող ISMA տեխնոլոգիայով կազմակերպվող էլեկտրոնային բառարանը ներկայացնում է իրական աշխարհի կոնցեպտները ներկայացնող կազմակերպչական միավորների՝ «կոնցեպտների» ցանցային կառուցվածք ունեցող էլեկտրոնային շտեմարան: «Կոնցեպտները» նույնականացվում են նրանց համապատասխանող իրական աշխարհի անվանումներով: Այս շտեմարանի ձևավորումն իրականացվում է լեզվագետի կողմից, որը պետք է հանդես գա իրական աշխարհի փորձագետի դերում: Նա պետք է վերլուծի իրական աշխարհի օբյեկտները, ձևավորի նրանց ընդհանուրը դասը և շտեմարան ներածի այդ դասին համապատասխանող կազմակերպչական միավորը՝ «կոնցեպտը»: Նա պետք է հատկորոշի շտեմարանի այդ միավորը՝ տալով նրա անվանումը բառարանում վարվող լեզուներով, և համապատասխանեցնի նրանց հարաբերությունները շտեմարանի այլ միավորների հետ:

Բառարանի հիմքում դրված գիտելիքների ցանցային շտեմարանը ձևավորելիս մեր կողմից առաջարկվում է ցանցի «կոնցեպտների» միջև տրվող կապերի կազմ, որոնք բաժանել ենք երկու խմբի. 1. կապեր, որոնք ներկայացնում են ցանցի «կոնցեպտների» միջև եղած իմաստային հարաբերությունները (դաս-տեսակ, մաս-ամբողջություն, ոլորտ, ժամանակահատված, պատճառ, հետևանք, աղբյուր, արդյունք, վրա, մեջ), 2. կապեր, որոնք ներկայացնում են ցանցի «կոնցեպտները» ներկայացնող անվանումների շարահյուսական հարաբերությունները (ստորոգյալ (ենթակա), ուղիղ խնդիր, հանգման խնդիր, փոխաբերական հանգման խնդիր, անջատման խնդիր, փոխաբերական անջատման խնդիր, վերաբերմունքի խնդիր, միջոցի խնդիր, հատկացուցիչ-հատկացյալ, ածական):

4. Բազմալեզու թարգմանական-բացատրական բառարանների կազմակերպումն ISMA տեխնոլոգիայով: Բազմալեզու բառարանների համար կոնցեպտը կարելի է կազմակերպել բազմատող աղյուսակի տեսքով, որի տողերում կարող են գրվել տվյալ կոնցեպտի անունը վարվող լեզուներից մեկով:

Թարգմանության ռեժիմում բառարանի մուտքին տրվում է թարգմանվող լեզվի անվանումը և այդ լեզվով թարգմանություն պահանջող բառը, ինչպես նաև բոլոր այն լեզուների անունները, որոնցով պահանջվում է ստանալ տվյալ բառի թարգմանությունը: Այսինքն ելքում ստացվում են տվյալ բառի թարգմանություններն իրենց հոմանիշներով միաժամանակ մի քանի լեզուներով:

5. Էքսպերտային համակարգ՝ ցանցային գիտելիքների բազայի սկզբունքով աշխատող բացատրական բառարանների ստեղծման համար: Նշված դասի բառարանների կազմման գործընթացում հիմնական խնդիրը կոնցեպտների կազմի ընտրությունն է: Ստորև առաջարկվում են էքսպերտային համակարգի աշխատանքի մեթոդ և սխեմա, որոնք թույլ կտան որոշակի տեսքով ձևայնացնել այդ կոնցեպտների ընտրման գործընթացը: Այս աշխատանքում ելնում ենք այն դրույթից, որ կոնցեպտների ձևավորման համար անհրաժեշտ հիմնական ինֆորմացիան առկա է դասական բացատրական բառարաններում: Ներկայումս նման բառարանները թվայնացված են և մատչելի են ծրագրային մշակման համար, որի արդյունքում թվայնացված բառարանը կարելի է ներկայացնել M երկչափ զանգվածով և նրան կից S սիմվոլային գրառումով (նկ.3):



Նկ. 3

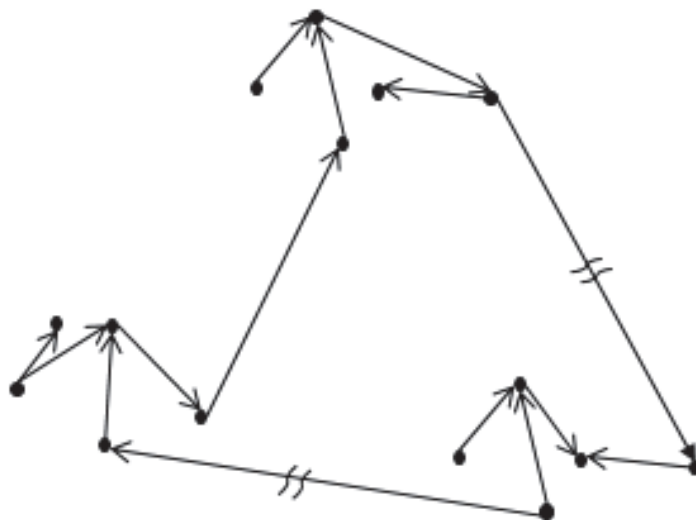
$M(i, 1)$ -ը պարունակում է հղում S գրառման այն ինդեքսին, որտեղից սկսվում է i -րդ բառարանային հոդվածի գրառումը: $M[i, 2]$ -ը $M[i, 1]$ -ում ներկայացվող բառի խոսքի - մասային ծածկագիրն է: Իսկ $M[i, 3]$ -ը պարունակում է հղում S գրառման այն ինդեքսին, որտեղից սկսվում է տվյալ բառարանային հոդվածի տեքստային նկարագրության գրառումը: Եթե բառը բազմիմաստ է, նրան համապատասխանում է $M[n, 3]$ մատրիցի մի քանի տող, որոնցում $M[i, 1]$ -ի արժեքները համընկնում են: Մեր կարծիքով վերը նշված էքսպերտային համակարգը (է<) պետք է կարողանա վերադասավորելով խմբավորել $M[n, 3]$ մատրիցի տողերն այնպես, որ այդ խմբերը բացատրական բառարանը ստեղծող լեզվագետին հուշեն այդ խմբին համապատասխանող կոնցեպտը: Նշված մատրիցի տողը կոնկրետ խմբի մեջ մտցնելու հայտանիշը այդ տողի $M[i, 3]$ դաշտով ներկայացվող $M[i, 1]$ բառի իմաստային մոտիկությունն է խմբի մեջ մտած այլ բառերի իմաստների հետ: Եթե խմբի մեջ մտնող $k1$ խմբի $M[k1, 3]$ դաշտով ներկայացված իմաստը խիստ մոտիկ է կամ համընկնում է մի այլ $k2$ տողի $M[k2, 3]$ դաշտով ներկայացված իմաստի հետ, ապա դա կարելի է մեկնաբանել, որ գիտելիքների բազայում ցանկալի է մտցնել կոնցեպտ, որի անունը կունենա 2 համարժեքներ (հոմանիշներ)՝ ներկայացված $M[k1,1]$, $M[k2,1]$ դաշտերով լրացված բառերով: Եթե խմբի մեջ մտնող տողերի երրորդ սյուններում ներկայացված իմաստները մոտիկ են, բայց չեն համընկնում, դա հուշում է, որ գիտելիքների բազայում պետք է ստեղծել կոնցեպտներ, որոնք պատկանում են միևնույն ենթադասին: Սա թույլ կտա լեզվաբանին ստրուկտուրավորել ԳԲ-ի սեմանտիկ ցանցը: Բառարանային հոդվածների վերը նշված մոտիկությունը գնահատելու համար առաջարկվում է օգտվել էլեկտրոնային գրադարանային համակարգերի ինֆորմացիոն որոնման ռեժիմներում օգտագործվող տեքստերի մոտիկության գնահատման հայտնի չափանիշներից [1,2]: Վերջիններս ներկայացվում են $R[i,j]$ մատրիցով, որի տարրը՝ $0 < r_{ij} < 1$, իսկ i -ն և j -ն համեմատվող տեքստերի համարներն են:

6. Էքսպերտային համակարգի աշխատանքի կազմակերպման սխեման: Այսպիսով, խնդիրը վերածվում է $R[i,j]$ մատրիցի միջոցով միմյանց իմաստով մոտիկ բառերի խմբերի որոնմանը: Մեր կողմից առաջարկվում է աշխատանքի սկզբում n - տարրանոց բառերի խումբը բաժանել երկու խմբի այնպես, որ բառերի ներխմբային կապերի միջին արժեքը մեծ լինի, քան միջխմբային բառերի կապերը, ապա տողերի և սյուների վերադասավորման հաշվին ստանալ նոր մատրից, որտեղ 1-ից մինչև m տողերով կներկայացվեն 1-ին խմբի բառերը, իսկ m -ից մինչև n -ով՝ երկրորդ խմբի բառերը: Ելնելով վերը նշված խմբավորումից՝ $R[i,j]$ ($n \times n$) սիմետրիկ մատրիցը կտրոհենք երկու՝ $R1$ և $R2$ մատրիցների: Հետագայում մատրիցը երկու մասի բաժանելու մոտեցումը կկիրառենք $R1$ և $R2$ մատրիցների նկատմամբ առանձին-առանձին: Այդ գործընթացը կշարունակվի այնքան, քանի դեռ ցանկացած խմբի ներխմբային կապերի միջին արժեքի և միջխմբային կապերի միջին արժեքի հարաբերությունը փոքր է տրված ε թվից: Վերը նշված մատրիցների երկատման համար առաջարկում ենք օգտագործել տաքսոնոմիայի տեսության հայտնի մեթոդը [3], որը կիրակականացնենք հետևյալ քայլերով.

1. փնտրում ենք $R[i,j]$ ($n \times n$) մատրիցի նվազագույն տարրը, ապա այն տեղադրում $R'[i,j]$ ($n \times n$) սիմետրիկ զրոյական մատրիցի մեջ, միևնույն ժամանակ $R[i,j]$ մատրիցի i -րդ տողի j -րդ սյան բոլոր տարրերին վերագրում ենք \max արժեք, որպեսզի դրանք հեռացնենք հետագա դիտարկումներից,

2. փնտրում ենք $R[i,j]$ մատրիցի մեջ նվազագույն R_{i_1,j_1} տարրը, որտեղ j_1 -ը $R'[i,j]$ մատրիցի ընթացիկ պահի n_2 զրոյական տարրերից մեկն է: Այս քայլը կրկնում ենք այնքան, քանի դեռ չեն դիտարկվել $R[i,j]$ մատրիցի բոլոր տարրերը, այսինքն, քանի դեռ նրա բոլոր տարրերին չեն վերագրվել \max արժեքներ: Հեշտ է նկատել, որ եթե $R'[i,j]$ մատրիցի տողերը դիտարկվեն որպես կետեր տարածության մեջ, իսկ նրա տարրերը՝ կապեր այդ կետերի միջև, ապա տարածության մեջ ստացված պատկերային գրաֆը չի ունենա ցիկլեր (նկ. 4):

3. Փնտրում ենք $R'[i,j]$ ($n \times n$) մատրիցի մեջ առավելագույն արժեք ունեցող տարրը և այն հեռացնում դիտարկումից: Քանի որ համապատասխան գրաֆը չուներ ցիկլեր, այդ գործողությունը գրաֆը բաժանում է երկու ինքնուրույն ենթագրաֆների, որոնք թույլ են տալիս, խմբավորելով այդ ենթագրաֆների հանգույցներին համապատասխանող տողերը, $R[i,j]$ մատրիցի մեջ բաժանել տրված $R[i,j]$ մատրիցը որոնելի $R1$ և $R2$ մատրիցների:



Նկ. 4

ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. **Кураленок И.Е., Некрестьянов И.С.** Оценка систем текстового поиска / МАИК <<Наука / Итерпериодика>> // Программирование.-2002.-№4.
2. **Барахнин В.Б., Федотов А.М.** Модели документального и фактографического поиска для интеллектуальных информационных систем // Электронный журнал Cloud of Science.-2014.-Т. 1, № 4.-С. 21.
3. **James M. Belk.** Thompson's group F: PhD Thesis / Cornell University, 2004.- 144 p.
URL: <https://www.imo.universite-paris-saclay.fr/~breuilla/Belk.pdf>

Օ.Վ. ՄԱՆՈՒԿՅԱՆ, Ա.Տ. ՄԱՆՈՒԿՅԱՆ, Յ.Ն. ՄԱՆՈՒԿՅԱՆ

ISMA ТЕХНОЛОГИЯ СОСТАВЛЕНИЯ ТОЛКОВЫХ СЛОВАРЕЙ

Предлагается использовать сетевую базу знаний для организации многоязычных толково-переводных словарей. Описывается организация многоязычных толково-переводных словарей по технологии ISMA и предлагается схема составления экспертной системы, предназначенной для оказания помощи лингвистам при создании толковых словарей, которая основана на использовании базы знаний сетевой структуры.

Ключевые слова: база знаний, словарная статья, толковый словарь, экспертная система, инверсное представление данных.

O.V. MANUKYAN, A.S. MANUKYAN, E.N. MANUKYAN

ISMA TECHNOLOGY FOR DEVELOPING DEFINITION DICTIONARIES

A network-based knowledge database is proposed as a modern way of organizing explanatory-translation dictionaries. The work describes the organization of multilingual explanatory-translation dictionaries by ISMA technology and produces a system of an expert organization scheme to support the creation of explanatory dictionaries based on the network knowledge database.

Keywords: knowledge database, dictionary article, explanatory dictionary, expert system, inverse data presentation.