

ՊԱՇՏՈՆԱԿԱՆ ԸՆԴԴԻՄԱԽՈՍԻ ԿԱՐԾԻՔ

Գևորգ Արմենի Պետրոսյանի

«Տեքստի ինքնատիպության աստիճանի որոշման բազմալեզու համակարգի մշակումը»

թեմայով, Ե.13.02 «Ավտոմատացման համակարգեր» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման ատենախոսության վերաբերյալ

Ատենախոսության թեմայի արդիականությունը և նպատակը

Գիտական և կրթական ոլորտներում տեղեկատվական հոսքերի և դրանց մշակման ինտենսիվության անշեղ աճը զգալիորեն մեծացրել են տեքստի ինքնատիպության աստիճանի որոշման ավտոմատացված համակարգերի մշակման անհրաժեշտությունը: Այս պայմաններում հատկապես արդիական է դարձել բազմալեզու միջավայրում տեքստային փոխառությունների հայտնաբերման և գիտական աշխատանքների ինքնատիպության աստիճանի որոշման ավտոմատացված համակարգերի մշակման անհրաժեշտությունը: Բազմալեզու միջավայրում տեքստի ինքնատիպության աստիճանի որոշման խնդիրը պահանջում է սկզբունքորեն նոր մեթոդներ՝ համեմատած տեքստային համապատասխանությունների դասական որոնման հետ:

Աշխատանքի առանձնահատուկ նշանակությունը կայանում է ոչ միայն մեկ լեզվի շրջանակներում ուղղակի համընկնումների հայտնաբերման, այլև հայերեն տեքստերում անգլերեն և ռուսերեն աղբյուրներից իմաստային համընկնումների հայտնաբերման վրա կենտրոնանալու մեջ: Այս մոտեցումը համապատասխանում է տեքստերի ինքնատիպության ստուգման և փոխառությունների հայտնաբերման ժամանակակից համակարգերին ներկայացվող պահանջներին և ուսումնասիրությունը դարձնում է խիստ արդիական:

Աշխատանքի նպատակը հայերեն տեքստի ինքնատիպության աստիճանի որոշման և դրանում պարունակվող միալեզու և միջլեզվական փոխառությունների հայտնաբերման ավտոմատացված համակարգի մշակումն է:

Ատենախոսության կառուցվածքը և բովանդակությունը

Ատենախոսությունը կառուցված է տրամաբանական հաջորդականությամբ և ներառում է ներածություն, հինգ գլուխ, եզրահանգում, օգտագործված գրականության ցանկ և հավելված:

Ներածությունում ձևակերպված են ատենախոսության թեմայի արդիականությունը, նպատակն ու խնդիրները, սահմանված են հետազոտության օբյեկտն ու առարկան, ինչպես

նաև թվարկված են պաշտպանության ներկայացվող հիմնական դրույթները, նշված են աշխատանքի գիտական նորույթն ու գործնական նշանակությունը:

Առաջին գլխում ներկայացված է տեքստային փոխառությունների հայտնաբերման և ինքնատիպության որոշման առկա մեթոդների ու համակարգերի համակողմանի վերլուծությունը: Մանրամասն ներկայացված են դրանց առավելություններն ու սահմանափակումները սակավաթիվ թվային ռեսուրսներով լեզուների դեպքում, և հիմնավորված է հայոց լեզվի համար մասնագիտացված մեթոդներ մշակելու անհրաժեշտությունը: Առանձնակի ուշադրություն է դարձվել միջլեզվական փոխառությունների հայտնաբերման առանձնահատկություններին, ինչպես նաև հայերենի՝ որպես սահմանափակ թվային ռեսուրսներով լեզվի խնդիրներին: Հատկապես ուշագրավ է միջլեզվական իմաստային համապատասխանեցման խնդրի համար խոր ուսուցման ժամանակակից մոդելների կիրառման կարևորությունը:

Երկրորդ գլխում ներկայացված են առաջարկվող ավտոմատացված համակարգի կառուցման ընդհանուր սկզբունքները: Մասնավորապես, ներկայացված են առաջարկվող համակարգի հայեցակարգային լուծումները, երկմակարդակ ճարտարապետությունը, ֆունկցիոնալ մոդելը և IDEFO տեխնոլոգիայով իրականացված նկարագրությունը, որոնք ապահովում են համակարգի կառուցվածքային ամբողջականությունն ու ընդլայնման հնարավորությունը: Նկարագրված է տեքստի մշակման բազմափուլ գործընթացը, որը ներառում է տեքստերի նախապատրաստման, հնարավոր սկզբնաղբյուրների ընտրության, հայերեն տեքստերի կառուցվածքային վերլուծության, մանրամասն վերլուծության և վերջնական արդյունքի ստեղծման փուլերը: Այս գլխում ներկայացված է նաև համակարգի ֆունկցիոնալ մոդելը, որն արտացոլում է համակարգի հիմնական բաղադրիչների փոխազդեցությունը:

Երրորդ գլխում ներկայացված են հեղինակի կողմից մշակված հիմնական մեթոդները: Առանձնապես արժեքավոր են.

- խոսքիմասային պիտակավորման վրա հիմնված թեկնածու տեքստերի որոնման մեթոդը,
- մարկովյան շղթաների վրա հիմնված միալեզու փոխառությունների որոնման մեթոդը,
- բազմալեզու տրանսֆորմերային մոդելի կիրառմամբ միջլեզվական փոխառությունների հայտնաբերման մեթոդը:

Նախադասությունների իմաստային համապատասխանեցման համար բազմալեզու տրանսֆորմերային մոդելի օգտագործումն արժանի է հատուկ ուշադրության: Այս գլխում ներկայացված են ընտրված LaBSE մոդելի կիրառումը, ինչպես նաև դրա վերապատրաստումը՝ օգտագործելով հատուկ պատրաստված տվյալների հավաքածուներ հայերեն-անգլերեն և հայերեն-ռուսերեն լեզվական զույգերի համար: Նկարագրված է վերապատրաստման տվյալների՝ զուգահեռ նախադասությունների, վերաձևակերպված զույգերի և բացասական օրինակների կառուցումը: Սակավաթիվ թվային ռեսուրսներով լեզուների համար նման տվյալների ստեղծումը բարդ և նշանակալի խնդիր է:

Չորրորդ գլխում ներկայացված են բազմակողմանի փորձարարական հետազոտություններ, կատարվել են համեմատական վերլուծություններ, գնահատվել են մշակված մեթոդների արդյունավետությունն ու ճշգրտությունը միջազգային ընդունված չափանիշների հիման վրա: Այս ամենը թույլ է տալիս գնահատել մեթոդները՝ հաշվի առնելով որոնման ամբողջականությունը, հայտնաբերման ճշգրտությունը և դասակարգման որակը: Համեմատական վերլուծությունների արդյունքները ցույց են տվել առաջարկվող լուծումների արդյունավետությունը:

Հինգերորդ գլխում ներկայացված է համակարգի ծրագրային իրականացումը, տվյալների բազայի կառուցվածքը, օգտագործողի վեբ ինտերֆեյսը և մշակված համակարգի համեմատությունը գոյություն ունեցող լուծումների հետ: Այս գլխի պարունակությունը հավաստում է, որ աշխատանքն ունի գործնական արդյունք՝ ավտոմատացված համակարգ, որը թույլ է տալիս բեռնել փաստաթղթեր, ստուգել դրանք, ցուցադրել գտնված հատվածները և ստեղծել հաշվետվություններ:

Եզրահանգման մեջ ներկայացված են ատենախոսության շրջանակներում ստացված հիմնական արդյունքները:

Ատենախոսության գիտական արդյունքների նորույթը և հավաստիությունը

Ատենախոսության հիմնական գիտական արդյունքներն ունեն ինքնուրույն բնույթ և պարունակում են գիտական նորույթ:

Հեղինակի կարևոր գիտական ձեռքբերումներից կարելի է առանձնացնել.

- հայերեն տեքստերում միջլեզվական փոխառությունների սկզբնաղբյուրների որոնման նոր մեթոդի մշակումը,
- մարկովյան շղթաների կիրառումը միալեզու փոխառությունների հայտնաբերման խնդրում,

- բազմալեզու LaBSE մոդելի վերապատրաստումը հայերեն-անգլերեն և հայերեն-ռուսերեն լեզվական զույգերի համար,

- հայերեն-անգլերեն և հայերեն-ռուսերեն վերաձևակերպված նախադասությունների հավաքածուների ստեղծումը,

- երկմակարդակ ավտոմատացված ճարտարապետության առաջարկումը, որը համատեղում է արագ նախնական որոնումն ու խոր իմաստաբանական վերլուծությունը:

Նշված արդյունքները զգալի ներդրում են բնական լեզվի մշակման և ավտոմատացման համակարգերի ոլորտում:

Գիտական դրույթների հավաստիությունը հաստատվում է մշակված մեթոդների մաթեմատիկական հիմնավորումներով և փորձարարական հետազոտություններով: Մշակված մեթոդների արդյունավետությունը գնահատվել է ինչպես թեստային, այնպես էլ իրական տվյալների հավաքածուների վրա՝ օգտագործելով մեթոդների գնահատման ընդունված չափանիշներ:

Գործնական նշանակությունը

Ատենախոսության արդյունքներն ունեն բարձր կիրառական արժեք:

Մշակված համակարգը հնարավորություն է տալիս.

- ավտոմատացնել գիտական աշխատանքների ինքնատիպության ստուգման գործընթացը,

- հայտնաբերել ոչ միայն բառացի, այլև վերաձևակերպված և թարգմանված փոխառությունները,

- աշխատել հայերեն, ռուսերեն և անգլերեն աղբյուրների հետ,

- կիրառվել Հայաստանի բուհերում, գիտահետազոտական կազմակերպություններում և գիտական հրատարակչություններում:

Համակարգի վեր իրականացման ընտրությունը, տվյալների կենտրոնացված պահպանումը և ասինխրոն մշակումների կիրառումը բարձրացնում են դրա արդյունավետությունն ու մասշտաբայնությունը:

Համակարգը պարզեցնում է փաստաթղթերի ստուգման գործընթացը, նվազեցնում ձեռքով վերլուծության աշխատատարությունը և օգտատիրոջը տրամադրում է հայտնաբերված համընկնումների վերաբերյալ հաշվետվություն:

Արդյունքների հիմնավորվածությունը

Ատենախոսությունում ներկայացված արդյունքները հիմնավորված են.

- ժամանակակից բնական լեզվի մշակման մեթոդների կիրառմամբ,
- տրանսֆորմերային ճարտարապետությունների օգտագործմամբ,
- փորձարարական բազմակողմանի ուսումնասիրություններով,
- միջազգային չափանիշներով իրականացված գնահատումներով,
- ծրագրային իրականացման և ներդրման փաստով:

Սա հնարավորություն է տալիս ներկայացված եզրակացությունները համարել հավաստի և գիտականորեն հիմնավորված:

Հարկ եմ համարում նշել ատենախոսության մեջ նկատված հետևյալ թերությունները:

1. Հեղինակը հիմնական ուշադրությունը կենտրոնացրել է հայերեն, ռուսերեն և անգլերեն լեզուների վրա: Ցանկալի կլիներ պարզաբանել, թե առաջարկված մեթոդներում ինչպիսի կարգաբերումներ են անհրաժեշտ այլ լեզուներին անցնելիս:
2. Աշխատանքում նպատակահարմար կլիներ առավել մանրամասն ներկայացնել մշակված համակարգի հաշվարկային բարդության և արտադրողականության վերլուծությունը մեծածավալ տվյալների մշակման դեպքում՝ մասնավորապես տեքստի ամբողջական ստուգման ժամանակային բնութագրերը բեռնումից մինչև հաշվետվության ստեղծում:
3. Ցանկալի կլիներ ներկայացնել առաջարկվող համակարգի ինտեգրման հնարավոր մեխանիզմները միջազգային գիտական շտեմարանների և տեղեկատվական համակարգերի հետ:
4. Ներկայացված փորձարարական արդյունքների ընկալելիության բարձրացման նպատակով, նպատակահարմար կլիներ ավելացնել վիճակագրական վստահության միջակայքին առնչվող լրացուցիչ մեկնաբանություններ:

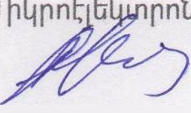
Եզրակացություն

Գևորգ Արմենի Պետրոսյանի «Տեքստի ինքնատիպության աստիճանի որոշման բազմալեզու համակարգի մշակումը» թեմայով ատենախոսությունը ավարտուն գիտահետազոտական աշխատանք է, որտեղ լուծված է արդիական գիտատեխնիկական խնդիր՝ հայերեն տեքստերի ինքնատիպության աստիճանի որոշման բազմալեզու ավտոմատացված համակարգի մշակումը:

Աշխատանքն առանձնանում է գիտական նորույթով, տեսական հիմնավորվածությամբ, կիրառական նշանակությամբ և ստացված արդյունքների հավաստիությամբ:

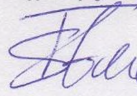
Հեղինակի հրատարակած 9 աշխատանքները և ատենախոսության սեղմագիրն արտահայտում են ատենախոսության հիմնական բովանդակությունը:

Կարծում եմ, որ չնայած վերը նշված դիտողություններին, ատենախոսությունը բավարարում է ՀՀ-ում գիտական աստիճանների շնորհման կարգի, թեկնածուական ատենախոսություններին ներկայացվող բոլոր պահանջներին՝ որպես գիտության տվյալ բնագավառում կիրառական կարևոր խնդրի լուծումն ապահովող գիտականորեն հիմնավորված տեխնիկական մշակում, իսկ հեղինակն արժանի է Ե.13.02 - «Ավտոմատացման համակարգեր» մասնագիտությամբ տեխնիկական գիտությունների թեկնածուի գիտական աստիճանի շնորհմանը:

Պաշտոնական ընդդիմախոս ՀԱՊՀ «Միկրոէլեկտրոնային սխեմաներ և համակարգեր»
ամբիոնի պրոֆեսոր, տ.գ.դ.  Ա.Գ.Հարությունյան

Տ.գ.դ. Ա.Գ. Հարությունյանի ստորագրությունը հաստատում եմ՝

ՀԱՊՀ-ի գիտական քարտուղար՝



Վ.Վ. Հովհաննիսյան

23.06.26թ.

