

Դուխիկ Խաչիկի Հովսեփյան

Հայցորդ

ՀԵՀ, SS և կիրառական մաթեմատիկայի ամբիոն

dukhik.hovsepyan@gmail.com

Ամփոփագիր

Ռեկացիոն տվյալների բազայի կառավարման համակարգերը վերջին 40 տարիների ընթացքում հանդիսացել են տվյալների պահպանման ստանդարտ մեխանիզմներ: Այս համակարգերը ստեղծվել են հստակ կազմակերպված և կանխատեսելի տեղեկատվության կառավարման համար՝ օգտագործելով՝ SQL հարցումների լեզու և հիմնվելով աղյուսակային ձևաչափով ներկայացումների վրա: Սակայն թվային տիրույթը կտրուկ փոփոխվել է: Ներկայումս տվյալները գնալով ավելի քառասյին և անկանխատեսելի են դառնում: Ներկայումս բազմաբովանդակ տեղեկատվական դաշտը գտնվում է այնպիսի ժամանակաշրջանում, երբ ոչ կառուցվածքային տվյալների ծավալը կտրուկ աճում է՝ սկսած բարդ մարդկային վարքագծի օրինաչափություններից մինչև բազմազան մուլտիմեդիա բովանդակություն, որոնք հաճախ դուրս են մնում ավանդական ռեկացիոն տվյալների բազաների արդյունավետ կիրառման շրջանակներից: Խնդիրը կայանում է նրանում, որ ավանդական տվյալների բազաների մեծ մասը նախատեսված է ճշգրիտ համընկնումների համար: Սակայն ժամանակակից ԱԲ համակարգերը պահանջում են տեղեկատվության որոնման շատ ավելի բարդ մակարդակ, որտեղ տեղեկատվությունը գտնվում է ոչ թե կոնկրետ հիմնաբառերի կամ տերմինների, այլ իրական իմաստի և համատեքստի հիման վրա:

Հետևաբար, վեկտորային տվյալների բազաների ի հայտ գալը շատ ավելին է, քան պարզապես տեխնիկական բարելավում. դա կազմակերպությունների համար իրենց տվյալների պահմանման և մշակման ենթակառուցվածքների հիմնովին վերափոխումն է: Ի տարբերություն ավանդական ճարտարապետությունների, վեկտորային տվյալների բազաները միավորում են կառուցվածքային և ոչ կառուցվածքային տվյալները մեկ համակարգում՝ ապահովելով կազմակերպություններին ավելի արագ ըմբռնումներ, իրապես արդյունավետության բարձրացում և բնական միջլեզվական հնարավորություն, ինչը հանդիսանում է ԱԲ-հավելվածների հիմնական կորիզը այսօր, ինչպես նաև հաջորդ տասնամյակների համար:

Այս հոդվածը հանգամանալից ուսումնասիրում է վեկտորային ներկայացումների տեսական հիմքերը, ընդգծում ռեկացիոն համակարգերի սահմանափակումները և ցույց տալիս, թե ինչպես

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

Են վեկտորային տվյալների բազաները վերաձևավորում տվյալների կառավարումը: Այն նաև վերլուծում է իրական կիրառության դեպքերը և փաստում, որ վեկտորային տվյալների բազաները ունեն առանցքային նշանակություն ԱԲ առաջնորդվող տեխնոլոգիական ապագայում:

Հիմնաբառեր. Վեկտորային տվյալների բազաներ, ռեկացիոն տվյալների բազաներ, ԱԲ տվյալների ենթակառուցվածք, իմաստային նմանություն, մոտարկվող ամենամոտ հարևանների որոնում:

Ներածություն

Տվյալների բազաները և մարդկության կողմից աշխարհի ներկայացման ձևը տվյալների միջոցով ձևավորվել են բազմաթիվ տարիների ընթացքում: Էդգար Ֆ. Կոդդը 1970 թ. ներկայացրեց ռեկացիոն մոդելը, որն իրենից ներկայացնում էր այնպիսի արստրակցիա, երբ տվյալները ստեղծվում և պահպանվում էին աղյուսակներում, դրանք հասանելի էին նկարագրողական լեզվի միջոցով և փոփոխվում էին երկրաչափական մեթոդներով: Այս մոդելը բազմաթիվ տարիների ընթացքում արդյունավետ միջոց է հանդիսացել տվյալների ներկայացման, պահպանման և մշակման համար՝ ապահովելով բիզնես հավելվածների լայն շրջանակ (ֆինանսական համակարգեր, էլ. առևտրի կայքեր, կորպորատիվ ռեսուրսների պլանավորում և այլն):

Տվյալների բնութագրիչները հսկայական փոփոխություններ են կրել: Ներկայումս տվյալների ստեղծման մեթոդներն ու գործիքները ստեղծում են հսկայական քանակությամբ նյութ, որը չի կարող ներկայացվել ռեկացիոն ձևաչափով: Օրինակ, սոցիալական ցանցերի հաղորդագրությունները, ռենտգենյան ճառագայթներով ստացված պատկերները, ապրանքային լուսանկարները, ծայնագրությունները, օգտատիրոջ ակտիվության մատյաններն ու նեյրոնային ցանցերի ակտիվացման վիճակները չեն կարող ներկայացվել աղյուսակային դաշտի արժեքների համընկնման միջոցով: Ժամանակակից տվյալների որոնումը դժվարացնում է ոչ թե ծավալը կամ արագությունը, այլ իմաստային վերլուծությունը:

Երբ օգտատերը լաբրադոր ռետրիվերի լուսանկար է ներբեռնում և խնդրում գտնել «նման պատկերներ», խելամիտ համակարգը պետք է վերադարձնի Գոլդեն ռետրիվերներ, Բիգլեր, գուցե Կոկեր-սպանիել, բայց ոչ մի սպորտային մեքենա, նույնիսկ եթե այդ մեքենայի մոտ գերակշռի օրինակ նույն գույնը կամ ֆայլի մետատվյալները: Այստեղ նմանությունը կարևորվում է, ոչ թե թվային պատահականությամբ. այլ իմաստային հատկություններով: Երկու շները իմաստային «մոտ» են այնպես, ինչպես շունն ու մեքենան իմաստով պարզապես «մոտ» չեն, և ցանկացած որոնման համակարգ, որն արժանի է իր անվանը, պետք է իմանա այս տարբերությունը:

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

Ռելացիոն տվյալների բազաները երբևէ չեն կառուցվել այսպիսի հարցերի համար: Նրանք բացարձակ լավ են կատարում այն խնդիրները, ինչի համար ստեղծվել են. գտնել տողը, որտեղ հաճախորդի կողմից հավասար է 4821-ի կամ վերադարձնել բոլոր պատվերները, որոնք կատարվել են երկու ամսաթվերի միջև: Սակայն, երբ խնդիր է դրվում չափել երկու պատկերի, երկու նախադասությունների կամ երկու գաղափարի հայեցակարգային հեռավորությունը, նրանք պատասխան չունեն: SQL-ում «բավականաչափ մոտ բառերով» հասկացություն գոյություն չունի: «Ինչ-որ կերպ շան նման» ինդեքս չկա:

Սա ոչ թե հեշտությամբ ուղղման ենթակա թերություն է, այլ ճարտարապետական եզրագիծ, սահման: Հենց այս սահմանի մոտ է, որ վեկտորային ներկայացումները հանդես են գալիս ոչ թե որպես ռելացիոն մոդելի ընդլայնում, այլ որպես այլընտրանքային հիմք, թե ինչպես պահպանենք, կառավարենք և մշակենք տվյալները: Վեկտորային տվյալների բազաները, կառուցված բարձրաչափ վեկտորային տարածության սկզբունքների հիման վրա, առաջարկում են տվյալների կառավարման այնպիսի մոտեցում, որը հնարավորություն է տալիս նմանության վրա հիմնված կատարել բարդ գործողություններ հատկապես ոչ կառուցվածքային տվյալների նկատմամբ [Han et al]:

Ռելացիոն տվյալների բազաների սահմանափակումներն իմաստային առաջադրանքների համար

Ռելացիոն մոդելը և նրա ենթադրությունները: Ռելացիոն մոդելի հիմքում ընկած են մի քանի հիմնական գաղափարներ, որոնք լավ աշխատել են 20-րդ դարի կեսի հաշվողական միջավայրերում, բայց այժմ ունեն շատ սահմանափակումներ ժամանակակից հավելվածների համար.

- ✓ այն ենթադրում է, որ տվյալները կարող են տարրալուծվել ատոմիկ արժեքների, որոնք կարող են համեմատվել միմյանց հետ, ինչպես ամբողջ թվեր, սիմվոլներ, ամսաթվեր և դրանց գումարային արժեքներ
- ✓ այն ենթադրում է, որ հավասարության պնդումները կամ միջակայքով տրված պայմանները կարող են ցույց տալ, թե ինչպես են գրառումները կապված միմյանց հետ
- ✓ այն պնդում է, որ սխեման, որը սահմանում է տվյալների կառուցվածքը, կանխորոշված է և համեմատաբար անփոփոխ:

Այս ենթադրությունները խախտվում են ոչ կառուցվածքային տվյալների նկատմամբ: Տրանսֆորմերային մոդելի կողմից տրված տեքստի հիման վրա ստեղծված 768-չափային վեկտորի ներկայացումը հնարավոր չէ համեմատել հավասարության/անհավասարության օպերատորների հետ: Նմանապես, «Գտիր ինձ 10 փաստաթղթեր, որոնք իմաստային առումով ամենաշատն են նման սրան» հարցումն ի վիճակի չէ բնականորեն արտահայտվել SQL-ով, և նույնիսկ ամենապարզ հնարավոր իրականացումը, որը ենթադրում է աղյուսակի յուրաքանչյուր

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

տողի վրա սկալյար արտադրյալի հաշվարկ, հանգեցնում է ամբողջ աղյուսակի սկանավորման և, հետևաբար, մասշտաբային տվյալների դեպքում կատարողականությունը իրատեսական չի [Abdellaoui S. et al]:

Իմաստաբանական ճեղք: Տեղեկատվության որոնման հիմնական մարտահրավերը իմաստաբանական ճեղքն է. համակարգի փաստացի տվյալների ներկայացման և օգտատերերի հարցումների մտածված իմաստների միջև անհամապատասխանությունը: Հիմնաբառային որոնումն ամենատարածված մեթոդն է ռելացիոն լրիվ-տեքստային որոնման համար (full-text search), որն ամբողջությամբ համեմատում է տարրեր, այլ ոչ թե իմաստները: Օրինակ, «մեքենաների ապահովագրություն» հարցումը չի ստանա «մեքենաների ծածկույթ» կամ «մեքենայի պաշտպանության քաղաքականություն» քննարկող փաստաթղթեր՝ առանց հստակ հոմանիշի ընդլայնման: Այս սահմանափակումները ավելի ցայտուն են բազմալեզու կայքերում: Հարաբերական տվյալների բազաները սովորաբար պահանջում են առանձին ինդեքսներ յուրաքանչյուր լեզվի համար՝ ստեղծելով խնդիրներ գլոբալ հավելվածների համար, ինչպեսիք են՝ առևտուրը և առողջապահությունը:

Արտադրողականության և ընդլայնելիության սահմանափակումները: Նույնիսկ, երբ ռելացիոն տվյալների բազաները փորձում են կրճատել իմաստային բացը՝ օգտագործելով ընդլայնումներ, ինչպիսիք են PostgreSQL-ի pg_trgm կամ tsvector-ի հետ լրիվ-տեքստային որոնումը, այս լուծումները դեռ բախվում են բնածին սահմանափակումների հետ: Trigram-ի վրա հիմնված համընկնումը, օրինակ, վատ է կատարում, երբ հարցումները կարճ են կամ ներառում են բարդ իմաստ: Նման ձևով, TF-IDF-ի նման վարկանիշավորման մեթոդները հիմնականում հենվում են մակերեսային բառային հաճախականության վրա: Այս մարտահրավերներն ավելի ցայտուն են դառնում մեծ մասշտաբների դեպքում, երբ բացակայում է արդյունավետ մատարկվող ամենամոտ հարևանի (ANN)-որոնման հնարավորությունը:

Վեկտորային ներկայացումներ. մաթեմատիկական հիմքեր

Ներկայացման ուսուցում: Ներկայացման ուսուցման հիմնական էությունն այն է, որ առարկաների իմաստային հատկությունները կարող են կոդավորվել որպես բազմաչափ վեկտորների երկրաչափական հատկություններ [Xiong]: Ուսուցված ներկայացումների ֆունկցիան արտապատկերում է ներածության առարկան (բառ, նախադասություն, պատկեր կամ օգտատիրոջ պրոֆիլ) որպես թվային արժեքներով վեկտոր, այնպես, որ նմանատիպ սեմանտիկ հատկություններ ունեցող առարկաները գտնվեն մոտ՝ ինչ-որ հեռավորության չափականության ներքո:

Word2Vec-ի և GloVe-ի նախնական բառի ներկայացման կառուցման մոտեցումները ցույց տվեցին, որ մեծ տեքստային բազաներում բառերի համատեղ հայտնվելու ձևանմուշները կարող են դուրս հանվել և պահպանվել կոմպակտ թվային վեկտորների միջոցով: Այս վեկտորները ի

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

վիճակի են պահպանել զարմանալիորեն հարուստ իմաստային հարաբերություններ: Օրինակ, «թագավոր», «տղամարդ», «կին» և «թագուհի» բառերը իմաստայնորեն պարզապես չեն պահվում սիմվոլային տեսքով, այլ կարող են նաև արտահայտվել վեկտորային տարածության մեջ հանրահաշվական հարաբերակցության միջոցով:

Համադրության և բազմաժանր ներկայացումներ: Ներդրումային ցանցերի ոլորտում տրանսֆորմեր ճարտարապետության վրա հիմնված մոդելների ներդրումը, ինչպիսիք են՝ BERT, RoBERTa, GPT-ն և դրանց հետագա տարբերակները, կտրուկ փոփոխություն բերեց տեքստի ներկայացման ձևաչափում: Ֆիքսված բառի ներկայացումներից կախվածության փոխարեն, այս մոդելները ստեղծում են համատեքստի վրա հիմնված ներկայացումներ: Հետագայում բազմաժանր մոդելները, ինչպիսին է CLIP-ը, ընդլայնում են այս հնարավորությունները, կառուցելով ըստ ժանրերի ընդհանուր տարածություններ, որտեղ պատկերներն ու դրանց տեքստային նկարագրությունները արտապատկերվում են մոտ կետերի: Այս ճարտարապետությունը հանգում է վերահաղորդագրություններով ամրացված սերնդավորման մեթոդի բարելավված հնարավորություններին (Retrieval-Augmented Generation - RAG):

Հեռավորության չափանիշներ և Նմանություն: Վեկտորային նմանությունն ամենից հաճախ չափվում է կշռված Jaccard-ի և կոսինուսային նմանության միջոցով: Կոսինուսային նմանությունը կանոնականացված է $[-1, 1]$ տիրույթում և չափում է վեկտորների միջև անկյունային հեռավորությունը, ինչը այն դարձնում է կայուն մեծությունների տարբերությունների նկատմամբ: Այլ չափանիշները ներառում են Էվկլիդյան (EuclideanL2) հեռավորությունը, ներքին արտադրյալի նմանությունը (MIPS-ում) և Հեմմինգի հեռավորությունը բինար ներկառուցումների համար: Վեկտորային տվյալների բազայի հիմնական գործողությունը մոտարկվող ամենամոտ հարևանների (ANN) որոնումն է. տրված է հարցման q վեկտոր, տվյալների բազա n վեկտորներով, պահանջվում է վերադարձնել k ամենամոտ վեկտորները q -ին ընտրված հեռավորության չափանիշի համաձայն [Yang et al]: ANN ալգորիթմները խաղարկում են գնահատման ճշգրտությունը՝ հատուկ ինդեքսային կառուցվածքների միջոցով, որպեսզի զգալիորեն բարելավեն հարցման ուշացումը [Gao et al.], [Oguri, Matsui]:

Վեկտորային Տվյալների Բազայի Ճարտարապետություն

Հիմնական Բաղադրիչներ: Վեկտորային տվյալների բազան նպատակային տվյալների կառավարման համակարգ է, որը նախատեսված է բազմաչափ վեկտորային ներկայացումների պահպանման, ինդեքսավորման և հարցումների համար: Նրա հիմնական բաղադրիչները ներառում են.

- ✓ վեկտորային պահպանման շերտ, որը պահպանում է վեկտորների ներկայացումները, այդ թվում նաև մետատվյալները,

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

- ✓ վեկտորային ինդեքս, որը հնարավորություն է տալիս արդյունավետ իրականացնել մոտարկվող ամենամոտ հարևանի (approximate nearest neighbor - ANN) որոնումները,
- ✓ հարցման շարժիչ (query engine), որն աշխատում է նմանության հարցումների հետ՝ անհրաժեշտության դեպքում կիրառելով մետատվյալների ֆիլտրեր;
- ✓ API շերտ, որը տրամադրում է CRUD գործողություններ և որոնման ինտերֆեյսներ:

Առաջատար կոմերցիոն վեկտորային տվյալների բազաներից են Faiss (Facebook AI Research), Milvus, Weaviate, Qdrant, Chroma, Pinecone և pgvector (PostgreSQL ընդլայնում): Գրաֆային ինդեքսները, ինչպիսին է HNSW (Hierarchical Navigable Small World)-ը, հաճախ նախընտրվում են ցածր ուշացման և բարձր հիշողության հավասարակշռության համար [Hu et al.]:

Ինդեքսավորման ռազմավարություններ: HNSW ալգորիթմը կառուցում է բազմաշերտ մոտիկության գրաֆ, որտեղ յուրաքանչյուր հանգույց կապված է իր մոտարկվող ամենամոտ հարևանների հետ: Որոնումը տեղի է ունենում գրաֆի վերին կոշտ շերտից դեպի ստորին շերտեր, ինչը ապահովում է ենթասահմանային հարցման ժամանակ բարձր ճշտության ցուցանիշ: HNSW-ն առաջարկում է արդյունավետ ճշտություն-ուշացում փոխհարաբերություն հիշողության մեջ պահվող ինդեքսների համար [Malkov, Yashunin,] և հանդիսանում է հիմնական ալգորիթմը բազմաթիվ կիրառություններում:

IVF (Inverted File Index) ինդեքսը բաժանում է վեկտորային տարածությունը Վոռոնոյի բջիջների՝ օգտագործելով կլաստերացում: Հարցման պահին որոնվում են միայն այն բջիջները, որոնք ամենամոտ են հարցման վեկտորին, ինչը նվազեցնում է իրական որոնման տարածքը: IVF-ն սովորաբար համակցվում է PQ (product quantization) մեթոդի հետ, որը վեկտորները սեղմում է՝ բաժանելով դրանք ենթատարածքներում, և ապա յուրաքանչյուր ենթատարածք կոդավորելով փոքր արժեքներով, ինչը հնարավորություն է տալիս 8–32 անգամ հիշողության խնայողություն ստանալ՝ համեմատաբար փոքր ճշտության անկումով:

Հիբրիդ որոնում և մեթադվյալների ֆիլտրում: Ժամանակակից վեկտորային տվյալների բազաները ավելի ու ավելի հաճախ աջակցում են հիբրիդային որոնմանը՝ խտացված վեկտորային նմանության և բանալի բառերի առանձին համապատասխանության (BM25) կամ կառուցվածքային մետատվյալների ֆիլտրացիայի համադրությամբ: Սա թույլ է տալիս, օրինակ, հարցումներ ձևակերպել այսպիսի ձևով՝ գտնել 10 առավել իմաստաբանական նմանությամբ ապրանքի նկարագրությունները, որոնք սահմանափակված են պահեստում առկա ապրանքներով և արժեն \$100-ից ցածր: Իմաստաբանական որոնումն ու բիզնես տրամաբանության պայմանների համադրումը արտադրական կիրառությունների համար անհրաժեշտ է և ներկայացնում է նշանակալի ճարտարապետական առավելություն ինչպես բուն ռեկացիոն, այնպես էլ բուն վեկտորային մոտեցումների նկատմամբ:

Այս ինտեգրված մոտեցման շնորհիվ օգտագործում են և՛ իմաստաբանական բացատրելիությունը, և՛ ճշգրիտ հատկանիշների համապատասխանությունը, ինչի արդյունքում

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

ստացվում են ավելի համապատասխան արդյունքներ [Xie Y. et al.]: Օրինակ՝ HNSW-էկոհամակարգերը հիանալի են արագ մոտարկվող ամենամոտ հարևանների հաշվարկում, սակայն մետատվյալների ֆիլտրացիայի ներդրումը կարող է ավելի արդյունավետ սահմանափակել որոնման տարածքը՝ հեռացնելով այն վեկտորները, որոնք չեն բավարարում որոշակի կատեգորիկ կամ թվային չափանիշներին:

Լեզվական անկախությունը որպես կառուցվածքային հատկություն

Իմաստային ներկայացումների կառուցման տարածությունների ընդհանրություն:

Ժամանակակից վեկտորային ներկառուցումների ամենատեսականորեն համոզիչ հատկությունը, հավանաբար, դրանց լեզվական անկախությունն է: Տասնյակ կամ հարյուրավոր լեզուներ ընդգրկող կորպուսների վրա ուսուցված բազմալեզու տրանսֆորմերային մոդելները սովորում են արտապատկերել տարբեր լեզուներից իմաստով համարժեք արտահայտությունները դեպի ընդհանուր ներկառուցման տարածության մոտ շրջաններ: Այս հատկությունը ձեռքով չի նախագծվել, այլ ինքնաբերաբար առաջանում է բաշխման հիպոթեզի հիման վրա. եթե երկու հասկացություններ տարբեր լեզուներում հանդիպում են նման համատեքստերում, դրանց վեկտորային ներկայացումները բազմաչափ տարածքում մոտակա կետերում են տեղավորվում [Hermann, Blunsom]:

mBERT, XLM-R և LaBSE մոդելները ցույց են տվել, որ նախադասության վեկտորային ներկայացումները կարող են պահպանել ամուր կապ լեզուների միջև: Օրինակ, հայերեն նախադասություն, անգլերեն հարցում և դրա ճապոներեն թարգմանությունը բոլորը կարող են ներկայացվել վեկտորներով, որոնք մոտ են միմյանց բարձրաչափ տարածքում, օրինակ՝ 768-կամ 1024-չափանի վեկտորներով: Անգլերեն «diabetes risk factors» արտահայտության և դրա հայերեն համարժեքի միջև նմանությունը բարձր կլինի, չնայած բառերը չեն կիսում ընդհանուր տառեր կամ ձևեր: Սա ապացուցում է, որ մոդելները գրանցում են իմաստը լեզուների միջև՝ կախված չլինելով մակերեսային բառային համընկնումներից:

Տեղեկատվության որոնման հեղուկանքներ: Այս կառուցվածքային հատկությունը էապես փոխում է բազմալեզու տեղեկատվության որոնման համակարգերի ճարտարապետությունը: Ռեկացիոն մոտեցման պարագայում բազմալեզու որոնումը պահանջում է.

- ✓ լեզվի հայտնաբերում,
- ✓ հարցումների կամ փաստաթղթերի մեքենայական թարգմանություն,
- ✓ լեզվի համար առանձին ինդեքսների ստեղծում,
- ✓ լեզվապաշտպան արդյունքների գնահատման նորմալացում:

Վեկտորային տվյալների բազայի դեպում այս քայլերից ոչ մեկը պարտադիր չէ: Հարցումը կողավորվում է բազմալեզու վեկտորային ներկայացումների մոդելի կողմից. ստացված վեկտորը համեմատվում է բոլոր փաստաթղթային վեկտորների հետ՝ անկախ դրանց աղբյուրի լեզվից:

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

Համակարգը կարող է հեշտությամբ ընդլայնվել նոր լեզուների աջակցման համար՝ առանց որևէ փոփոխության տվյալների բազայի սխեմայում:

Փորձարարական ապացույց: XTREME և XGLUE բենչմարկները համակարգված կերպով գնահատում են լեզուների միջ-կրոսային փոխանցումը: Արդյունքները հետևողականորեն ցույց են տալիս, որ բազմալեզու ներկառուցումներով խիտ որոնումը գերազանցում է նոսր բառային մեթոդներին (BM25, TF-IDF) 20-40 կետով: Ավելին, zero-shot լեզուների միջ-կրոսային փոխանցումը ցույց է տալիս, թե որքան հետևողական է ներկառուցման տարածությունների սեմանտիկ կառուցվածքն ամբողջ լեզուներում:

XTREME և XGLUE չափորոշիչները համակարգված կերպով գնահատում են բազմալեզու անցումները տարբեր խնդիրներում, ինչպիսիք են՝ բազմալեզու փաստաթղթերի որոնումը, հարց-պատասխան համակարգերը և անունների ճանաչումը: Այս չափորոշիչների արդյունքները միանշանակ ցույց են տալիս, որ խտացված վեկտորային որոնումը բազմալեզու վեկտորային ներկայացումների միջոցով զգալիորեն գերազանցում է առանձին բառային մեթոդներին (BM25, TF-IDF) հաճախ 20–40 միավորով միջին ճշգրտությամբ ցուցանիշով, նույնիսկ այն դեպքում, երբ առանձին բառային մեթոդը ունի ոսկե-ստանդարտ թարգմանությունների հասանելիություն:

Ավելին, զրոյական օրինակով բազմալեզու անցումը, երբ մոդելը պատրաստված է միայն անգլերեն տվյալների վրա, բայց կարող է հաջողությամբ կատարել առաջադրանքներ ամբողջությամբ նոր լեզուներում, ցույց է տալիս, թե որքան կայուն է վեկտորային իմաստաբանական կառուցվածքը տարբեր լեզուների միջև: Այս կարողությունը բխում է այն փաստից, որով իմաստը երկրաչափորեն կոդավորվում է վեկտորային տարածքում՝ արտացոլելով իրական իմաստային հարաբերությունները՝ անկախ որևէ կոնկրետ ուսուցման կարգավորումից:

Վեկտորային Տվյալների Բազաների Իրական Կիրառությունները

Իմաստային որոնում և ձեռնարկային գիտելիքի կառավարում: Ձեռնարկային որոնումն այն կիրառություններից մեկն է, որտեղ վեկտորային տվյալների բազաները անմիջապես ցույց են տալիս իրենց բարձր արժեքը: Ավանդական, բանալի բառերի վրա հիմնված ձեռնարկային որոնման համակարգերը, որոնք ինդեքսավորված են Elasticsearch կամ Solr միջոցով, պահանջում են, որ օգտատերը կազմի հարցումները՝ օգտագործելով փաստաթղթերում առկա հստակ բառապաշարը [Löffler et al.]: Վեկտորային տվյալների բազաների հիման վրա կառուցված իմաստաբանական որոնման համակարգերը թույլ են տալիս աշխատակիցներին հարցումներ ներկայացնել բնական լեզվով, հայտնաբերել համապատասխան փաստաթղթեր, նույնիսկ այլ տերմինաբանություն օգտագործելով, որոնել տեղեկատվություն տարբեր փաստաթղթային ձևաչափերից (PDF-ներ, ներկայացումներ, էլ. նամակներ, տոմսեր) և տարբեր լեզուներից:

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

Ընկերությունները, ինչպիսիք են Notion, Atlassian և Slack, արդեն կիրառել են վեկտորային տվյալների վրա հիմնված իմաստաբանական որոնման հնարավորությունները և հայտնում են, որ զգալիորեն բարձրացել է արդյունքների համապատասխանությունը և կրճատվել է տեղեկություն ստանալու ժամանակը ձեռնարկներով աշխատող անձնակազմի համար:

Վերահաղորդագրություններով ամրացված սերնդավորման մեթոդ (Retrieval-Augmented Generation RAG): Ներկայումս RAG ճարտարապետությունն ընդլայնված կիրառություն ունի և համատեղում է մեծածավալ լեզվական մոդելները (LLMs) վեկտորային տվյալների բազայի որոնման հետ: RAG համակարգում օգտատիրոջ հարցումը կոդավորվում է որպես վեկտոր, ապա վեկտորային տվյալների բազան փաստաթղթային կորպուսից վերադարձնում է իմաստով ամենամոտ հատվածները: Այս հատվածները նախադրվում են LLM մոդելի հրահանգին որպես կոնտեքստ, և LLM մոդելը ստեղծում է փաստագրված, հիմնավորված և վստահելի պատասխան: RAG ճարտարապետությունները դարձել են ձեռնարկություններում LLM մոդելների հիմնական ներդրման մոտեցումը [Wang et al.], փոխարինելով այն անթույլատրելի մեթոդը, երբ մոդելները fine-tuning-ի միջոցով հարմարեցվում էին հատուկ տվյալների վրա: Վեկտորային տվյալների բազաները հանդիսանում են այս ճարտարապետության առանցքային ենթակառուցվածքային բաղադրիչը, երբ ռեկացիոն համակարգերը չեն ապահովում անհրաժեշտ ANN որոնումը ընդունելի ուշացումով և ծախսով: 2023–2025 թվականներին RAG լուծումների աճը հանգեցրել է Pinecone, Weaviate և Chroma պլատֆորմներում ակտիվ ներդրումների նշանակալի ավելացմանը:

Առաջարկության համակարգեր: Առաջարկային համակարգերը, լինի դա համագործակցային ֆիլտրում, թե բովանդակության վրա հիմնված մեթոդ, միշտ աշխատել են օգտատիրոջ և փնտրվող տարրի ներկայացումների հետ, որոնք բնականաբար վեկտորային բնույթ ունեն: Օրինակ՝ մատրիցայի ֆակտորացումը գեներացնում է թաքնված գործոնների վեկտորներ, մինչդեռ ներդրման համագործակցային ֆիլտրացիան և two-tower մոդելները համակարգին ուսուցանում են ներկայացումներ (embeddings), որոնք արտացոլում են օգտատիրոջ նախասիրություններն ու տարրի հատկանիշները [Wang et al.]: Վեկտորային տվյալների բազաները հնարավորություն են տալիս մեծամասշտաբային օգտագործել այս ներկայացումները: Երբ օգտատերը խնդրում է առաջարկություններ տրամադրել, նրա խնդրանքի վեկտորային ներկայացումը կարող է համեմատվել տարրերի ներկայացումների հետ միլիվայրկյանների ընթացքում և արագ ու լավագույնս դուրս բերել համընկնումները:

Իրական աշխարհի պլատֆորմները ցույց են տալիս այս մոտեցման արդյունավետությունը. Spotify երաժշտության համար, Pinterest՝ պատկերների հայտնաբերման և Airbnb՝ ցուցակների դեպքում բոլորն օգտվում են վեկտորային որոնման ենթակառուցվածքից՝ անձնականացված արդյունքներ իրական ժամանակում ապահովելու համար: Նրանց փորձը ընդգծում է, որ հատուկ վեկտորային պահոցները տրամադրում են զգալիորեն ավելի լավ կատարողականություն և

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

ընդլայնելիություն, քան փորձելը նման գործառույթները տեղավորել ավանդական, ընդհանուր նպատակ ունեցող տվյալների բազաներում:

Կենսաբժշկական և գիտական կիրառություններ: Վեկտորային տվյալների բազաները այժմ լայնորեն օգտագործվում են գիտական ոլորտներում, որտեղ տվյալները իրենց բնույթով բազմաչափ են: Դրանք ներառում են՝ սպիտակուցների կառուցվածքի ներկայացումներ, ինչպիսիք են ESM-2 ներկայացումները ամինաթթվի հաջորդականությունների համար, մոլեկուլային մատնահետքերը դեղագործական հայտնագործությունների մեջ, մեծածավալ աստղաբանական դիտարկումներ և նեյրոիմիջավորման առանձնահատկությունների տվյալներ: Այս ոլորտներում հետազոտողներին հաճախ անհրաժեշտ է որոնել իրադարձությունների միլիարդավոր կետերում՝ գտնելու իմաստով մոտ նմանություններ: Օրինակ նրանք կարող են փնտրել նման կառուցվածքով սպիտակուցներ, հարակից հատկություններ ունեցող քիմիական միացություններ կամ ընդհանուր հատկություններ ունեցող աստղեր: Այսպիսի նմանության որոնումները չափազանց դժվար են իրականացնել ավանդական ռեկացիոն համակարգերով, որոնք նախատեսված չեն այս տեսակի առաջադրանքների համար [Ol'ha et al.]:

Աչքի ընկնող օրինակ կարող է հանդիսանալ AlphaFold Protein Structure Database-ը, որը պարունակում է ավելի քան 200 միլիոն կանխատեսված սպիտակուցների կառուցվածքներ: Այսպիսի մասշտաբով տվյալների հետ աշխատելը պահանջում է բարձր արդյունավետությամբ որոնման մեթոդներ: Վեկտորային տվյալների բազաները դարձնում են դա հնարավոր՝ թույլ տալով մեծածավալ տվյալներից նմանության արագ որոնում: Ռեկացիոն մոդելները անհամեմատ դժվարությամբ են կառավարում այսպիսի աշխատանքային հոսքերը և շատ դեպքերում անիրագործելի են գիտական մեծածավալ վերլուծությունների համար:

Բանկային կիրառություններ: Վեկտորային տվյալների բազաները գնալով ավելի են կիրառվում ֆինանսական ոլորտում, որտեղ տվյալները բարդ են, բազմաչափ և անընդհատ փոփոխվում են: Բանկերը և ֆինանսական ընկերությունները մշակում են հսկայական ծավալի գործարքային գրառումներ, հաճախորդների վարքի ձևանմուշներ, վարկային պատմություններ:

Վեկտորային տվյալների բազաները հնարավորություն են տալիս իրական ժամանակում ստացվող նոր գործարքները զուգահեռել հազարավոր միլիոնավոր պատմական վեկտորների հետ՝ գրեթե անմիջապես հայտնաբերելու անսովոր կամ հնարավոր կասկածելի գործունեությունը: Սա այնպիսի խնդիր է, որը ավանդական հարաբերական համակարգերում կլինի չափազանց դանդաղ կամ անիրագործելի: Գողության կանխարգելումից դուրս, վեկտորային ներկայացումները նաև բարելավում են անձնականացված առաջարկությունները և ռիսկերի գնահատումները, օգնելով բանկերին առաջարկել հատուկ ֆինանսական ապրանքներ և ընդունել ավելի հիմնավորված վարկավորման որոշումներ:

Ընդհանուր առմամբ, այս հնարավորությունները ցույց են տալիս, թե ինչպես են վեկտորային տվյալների բազաները ֆինանսական ոլորտին տրամադրում ընդլայնելի և բարձր կատարողականությամբ գործիքներ: Դրանք թույլ են տալիս հաստատություններին իրական

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

Ժամանակում վերլուծել բարդ տվյալների հավաքածուներ, հայտնաբերել թաքնված ձևանմուշներ և ընդունել խելամիտ, արագ որոշումներ, այսինքն կատարել այնպիսի առաջադրանքներ, որոնք ավանդական տվյալների բազաների տեխնոլոգիայով կոժվարանային կամ անիրագործելի կլինեին:

Ներկայիս մարտահրավերներ և հեղազոտության ուղղություններ

Հերկողականություն, կայունություն և ACID երաշխիքներ: Վեկտորային տվյալների բազաների նկատմամբ արդարացի մտահոգություն է կազմում հատկապես ձեռնարկություններում և կարգավորվող միջավայրերում դրանց գործարքային երաշխիքների համեմատական անվիճելիությունը: Շատ վեկտորային բազաներ սկզբում նախագծվել են ընթերցման ծանրաբեռնվածության, մոտավոր որոնման աշխատանքային հոսքերի համար և չեն իրականացնում ամբողջական ACID (Atomicity, Consistency, Isolation, Durability) սեմանտիկա: Գրաֆների վրա հիմնված ինդեքսներում (օրինակ՝ HNSW) հաճախակի թարմացումների առկայության դեպքում ինդեքսի հուսալիությունը դեռևս չլուծված ինժեներական մարտահրավեր է:

Վերջերս իրականացված աշխատանքները, ինչպիսիք են՝ ինդեքսների հոսքային թարմացումները, վեկտորային պահոցների համար լոգավորումը և վեկտորային որոնման ինտեգրումը գործարքային համակարգերում (օրինակ՝ pgvector PostgreSQL-ում), մասնակիորեն լուծում են այս խնդիրները: Սակայն այն հավելվածների համար, որոնք պահանջում են խստագույն համատեղելիության երաշխիքներ, ինչպիսիք են ֆինանսական գրառումները կամ բժշկական ֆայլերը, գործնական ճարտարապետությունը շարունակում է մնալ հիբրիդային՝ համատեղ օգտագործելով OLTP համակարգի և վեկտորային որոնման շերտի համադրությունը:

Ներկայացումների որակ և տիրույթի հարմարվողականություն: Վեկտորային ստացման որակը հիմնականում որոշվում է հիմքում ընկած ներկայացումների կառուցման մոդելի կատարողականով: Ընդհանուր նպատակային բազմալեզու մոդելները հաճախ դժվարությունների են հանդիպում, երբ կիրառվում են հատուկ ոլորտներում, օրինակ՝ իրավաբանական, կլինիկական կամ գիտական համատեքստերում, քանի որ դրանց ուսուցման տվյալները չեն արտացոլում այդ ոլորտների հատուկ բառապաշարն ու իմաստաբանական կառուցվածքը: Մոդելի բարելավումը տվյալ ոլորտին հատուկ տվյալների վրա կարող է նշանակալիորեն լավացնել արդյունքները, սակայն այս մոտեցումը պահանջում է մեծ ծավալով նշագրված տվյալներ, զգալի հաշվարկային ռեսուրսներ և մոդելների շարունակական և զգուշավոր կառավարմամբ ուղեկցվող գործընթաց:

Մեկ այլ մարտահրավեր է հատուկ ոլորտներում բազմաչափ վեկտորների միջև իրական նմանությունը և տարբերությունը ճշգրիտ արձանագրելը: Սա մի խնդիր է, որը շարունակում է մնալ ակտիվ հետազոտական ոլորտում: Ընթացիկ ջանքերը կենտրոնանում են մի քանի խոստումնալից ուղղությունների վրա, ինչպիսիք են ցածր ռեսուրսներով տիրույթների

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

հարմարվելու մեթոդները contrastive learning, instruction-tuned embedding մոդելներ, որոնցից են օրինակ E5-Instruct և GTE-Qwen, և որոնման վրա կենտրոնացված նախնական ուսուցման (pre-training) նպատակները, որոնք նախատեսված են վեկտորային ներկայացումների որակը բարելավելու համար հատուկ որոնողական առաջադրանքներում: Այս մոտեցումների նպատակն է վեկտորային ներկայացումները դարձնել ավելի հստակ և արդյունավետ՝ ոլորտին հատուկ որոնողական կիրառումների համար:

Բացադրելիություն և վերահսկելիություն: Խիտ վեկտորային ստացման հիմնական մարտահրավերն այն է, որ այն հիմնականում ավելի քիչ բացատրելի է, քան ավանդական բառային որոնման համակարգերը: Ստանդարտ SQL հարցումում տվյալների ստացման չափանիշները հստակ են: Ի հակադրություն, մոտարկվող ամենամոտ հարևանների (ANN) որոնումները կատարվում են բարձրաչափ վեկտորային տարածություններում, և հաճախ անհասկանալի է, թե ինչու է կոնկրետ արդյունք վերադարձվել: Թափանցիկության այս պակասը կարող է լուրջ խնդիր լինել կարգավորվող ոլորտներում, որտեղ կարևոր են պատասխանատվությունն ու հետազոտելիությունը: Այս սահմանափակումները հաղթահարելու նպատակով մշակվել են հիբրիդային որոնման համակարգեր, որոնք համատեղում են խտացված ներկայացումները և ցրված բանալի բառերի վրա հիմնված ազդակները: Այս համակարգերը ապահովում են իմաստաբանական ճկունության և բացատրելիության միջև հավասարակշռություն: Ներկա պահին հետզհետե ավելի շատ են օգտագործվում նաև պատկանելիության մեթոդները, որոնք նպատակ ունեն ընդգծել այն կոնկրետ չափորոշիչները կամ բաղադրիչները, որոնք առավել ազդեցիկ են նմանության գնահատման վրա: Բացահայտելով, թե վեկտորի որ տարրերն են նպաստել համընկման ստեղծմանը, այս տեխնիկաները դարձնում են խտացված որոնման արդյունքները ավելի հասկանալի և գործնական ինչպես օգտատերերի, այնպես էլ վերահսկողների համար:

Եզրակացություն

Ռեւալացիոն տվյալների բազաները ծախսողվում են ոչ թե վատ նախագծված կամ իրականացված լինելու պատճառով, այլ դրանք բախվում են այնպիսի խնդիրների, որոնց համար երբևէ նախատեսված չէին: Իմաստային նմանության որոնման, տարբեր լեզուներով ստացման, բազմաժանր համընկնման, ԱԲ-ով առաջնորդվող գիտելիքների հայտնաբերման և այլ նմանատիպ առաջադրանքները դուրս են ռեւալացիոն մոդելի հնարավորություններից:

Վեկտորային տվյալների բազաները լրացնում են այս բացը ոչ թե գոյություն ունեցող համակարգերի բարելավումների, այլ տվյալների ստացումը երկրաչափական տերմիններով ամբողջությամբ վերափոխելու միջոցով:

Վեկտորային որոնման կարևորությունը շարունակապես աճում է: Մեծ լեզվային մոդելների, վերահաղորդագրություններով ամրացված սերնդավորման մեթոդի (RAG) և ԱԲ հավելվածների զարգացումը ստեղծել է իմաստային որոնման գործառույթների մեծ անհրաժեշտություն, որոնք

ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ

ավանդական ռեկացիոն տվյալների բազաները չեն կարող ապահովել: Վեկտորային տվյալների բազաները դարձել են նպատակային լուծում՝ բավարարելու այս պահանջը, և դրանց զարգացումը, մասնավորապես ընդլայնելիության, գործարքային երաշխիքների և գործառնական գործիքների առումով, բավականին արագ է ընթանում:

Այսպիսով, վեկտորային տվյալների բազաները պատրաստ են դառնալ խելամիտ տվյալների ստացման գերիշխող մեթոդ: Սա ռեկացիոն տվյալների բազաները հնացած և անօգտավետ չի դարձնում. նրանք շարունակելու են ծառայել որպես կառուցվածքային, գործարքային տեղեկատվության հեղինակային պահեստ:

Ավելի շատ այս մոտեցումը շեշտում է հետերոգեններ ճարտարապետության կարևորությունը, որտեղ վեկտորային բազաները առաջնորդում են այն գործառույթները, որոնք հիմնված են իմաստաբանական նմանությամբ, և այս կատեգորիան կշարունակի աճել ԱԲ հավելվածների տարածման և ոչ կառուցվածքային տվյալների ծավալների աննախադեպ աճի ֆոնին:

Գրականության ցանկ

1. Xiong B. Geometric Relational Embeddings. arXiv (Cornell University). 2024, 177 p., <https://doi.org/10.48550/arXiv.2409.15369>,
2. Abdellaoui S., Abbaoui W., Meziane L., Bhiri B. E., Ziti S. SQL and NoSQL Databases: A Comparative Study with Perspectives on IA-Based Migration Approach. 72. 2025, <https://doi.org/10.3390/engproc2025112072>,
3. Gao J., Gou Y., Xu Y., Yang Y., Cheng L., Wong R. C.-W. Practical and Asymptotically Optimal Quantization of High-Dimensional Vectors in Euclidean Space for Approximate Nearest Neighbor Search. arXiv (Cornell University). 2024. <https://doi.org/10.48550/arxiv.2409.09913>,
4. Han Y., Liu C., Wang P. A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge, arXiv (Cornell University)., 2023, <https://doi.org/10.48550/arXiv.2310.11703>,
5. Hermann K. M., Blunsom P. Multilingual Models for Compositional Distributed Semantics. ACL, 2014. pp. 58–68 , <https://doi.org/10.3115/v1/p14-1006>,
6. Hu G., Cai S., Dinh T. T. A., Xie Z., Yue C., Chen G., Ooi, B. C. HAKES: Scalable Vector Database for Embedding Search Service. PVLDB, 18(9), 2025, pp. 3049-3062, <https://doi.org/10.48550/ARXIV.2505.12524>,
7. Löffler F., Wesp V., König-Ries B., Klan F. Dataset search in biodiversity research: Do metadata in data repositories reflect scholarly information needs? PLoS ONE, 16(3), 2021, <https://doi.org/10.1371/journal.pone.0246099>,

8. Malkov Yu. A., Yashunin, D. A. Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(4), pp. 824-836, 2018,
<https://doi.org/10.1109/tpami.2018.2889473>,
9. Oguri Y., Matsui Y. Theoretical and Empirical Analysis of Adaptive Entry Point Selection for Graph-based Approximate Nearest Neighbor Search. arXiv (Cornell University). 2024.
<https://doi.org/10.48550/arxiv.2402.04713>,
10. Ol'ha J., Slanináková T., Gendiar M., Antol M., Dohnal,V. Learned Indexing in Proteins: Extended Work on Substituting Complex Distance Calculations with Embedding and Clustering Techniques. arXiv (Cornell University). 2022. <https://doi.org/10.48550/arxiv.2208.08910>,
11. Wang T., Brovman Y. M., Madhvanath, S. Personalized Embedding-based e-Commerce Recommendations at eBay. arXiv (Cornell University). 2021.
<http://arxiv.org/abs/2102.06156> , 30.03.2026,
12. Wang Y., Hernandez A. G., Kyslyi R., Kersting, N. Evaluating Quality of Answers for Retrieval-Augmented Generation: A Strong LLM Is All You Need. arXiv (Cornell University). 2024,
<https://doi.org/10.48550/arxiv.2406.18064>,
13. Xie Y., Hou, X.,Zhao, Y., Wang S., Chen K., Wang H.. Toward Understanding Bugs in Vector Database Management Systems. arXiv (Cornell University). 2025,
<https://doi.org/10.48550/arxiv.2506.02617>,
14. Yang M., Li W., Jin J., Zhong X., Wang X., Shen Z., Jia,W., Wang W. Effective and General Distance Computation for Approximate Nearest Neighbor Search. arXiv (Cornell University). 2024, <https://doi.org/10.48550/arxiv.2404.16322>,
15. Zhitomirsky-Geffet M., Dagan I. Bootstrapping Distributional Feature Vector Quality. Computational Linguistics, 35(3), 2009. pp 435-461,
<https://doi.org/10.1162/coli.08-032-r1-06-96>.

THE PARADIGM SHIFT TOWARD SEMANTIC SIMILARITY AND VECTOR DATA STORES

Dukhik Hovsepyan

PhD Researcher,

EUA, Chair of Information Technologies and Applied Mathematics

dukhik.hovsepyan@gmail.com

Abstract

Relational Database Management Systems have been the standard for data storage for the last 40 years. These systems were built to excel at managing highly organized and predictable information using Structured Query Language and rigid table formats. However, the digital landscape has shifted dramatically. Currently, the data we produce is increasingly chaotic and unpredictable. We have seen an explosion of messy, unstructured information ranging from complex human behavioral patterns to diverse multimedia that simply hits a dead end when faced with the constraints of a traditional relational database setup. The core of this issue is that most traditional databases are designed for exact matches. However, modern AI systems require a much more sophisticated level of retrieval, where information is found based on its actual meaning and context, rather than just specific keywords or terms. Therefore, the rise of vector databases represents far more than a simple technical upgrade; it is a fundamental reimagining of how organizations build their data infrastructure. Unlike traditional architectures, vector databases unify structured and unstructured data into a single system, providing organizations with faster insights, real efficiency gains, and a natural cross-language capability that positions them as the core foundation of AI-driven applications over the next decade.

This article explores the theoretical foundations of vector representations, highlights the limitations of relational systems, and demonstrates how vector databases are reshaping data management. It also discusses real-world use cases and argues that vector databases are likely to play a central role in the future of AI-driven technology.

Keywords: Vector databases, relational databases, AI data infrastructure, semantic similarity, approximate nearest neighbor search.

Ներկայացվել է՝ 05.04.2026թ.

Ուղարկվել է գրախոսման՝ 07.05.2026թ.