

AXBOROT QIDIRUV JARAYONIDA ONTOLOGIK BILIMLAR BAZASINING SAMARADORLIGINI TAHLIL QILISH

Babajanov Umar Usmonovich,
o'qituvchisi

umarbek.babajanov114@gmail.com

Abu Rayhon Beruniy nomidagi Urganch davlat universitetli

Annotatsiya: Ushbu maqolada axborot qidiruv tizimlarida ontologik bilimlar bazasining o'rnini tahlil qilinadi. An'anaviy qidiruvdan farqli ravishda semantik qidiruv modellarining afzalliklari, xususan, o'zbek tili NLP tizimlarida ontologiyalarni joriy etish masalalari yoritilgan.

Kalit so'zlar: *axborot qidiruvi, ontologiya, semantik qidiruv, O'zbek NLP, kompyuter lingvistikasi, korpus.*

Abstract: This article analyzes the role of ontological knowledge bases in information retrieval systems. Unlike traditional search, the advantages of semantic search models, particularly the implementation of ontologies in Uzbek NLP systems, are highlighted.

Keywords: *information retrieval, ontology, semantic search, Uzbek NLP, computational linguistics, corpus.*

Bugungi kunda axborot-kommunikatsiya texnologiyalarining mislsiz darajada rivojlanishi, kompyuter tarmoqlari quvvatining oshishi va internet tarmog'ining global qamrovi natijasida insoniyat "axborot portlashi" deb nomlanuvchi davrni boshdan kechirmoqda. Bunday ulkan hajmdagi "Katta ma'lumotlar" dengizida foydalanuvchilar uchun o'ziga kerakli, aniq va ishonchli axborotni qisqa vaqt ichida topish masalasi tobora murakkablashib bormoqda. Axborot qidiruv tizimlari aynan shu ehtiyojni qondirish maqsadida yaratilgan bo'lib, ular uzoq yillar davomida insoniyatning raqamli ma'lumotlar bilan ishlashdagi asosiy intellektual ko'prigi bo'lib xizmat qilib kelmoqda. Biroq, an'anaviy axborot qidiruv tizimlarining aksariyati shu paytgacha asosan kalit so'zlarga asoslangan mantiqiy yoki statistik modellarga tayanadi. Bu modellar garchi arxitekturasi



jihatidan sodda va hisoblash quvvati borasida tezkor bo'lsa-da, inson tilining o'ta murakkab tabiati, uning yashirin semantik va pragmatik xususiyatlarini inobatga olmagan uchun ko'p hollarda foydalanuvchi so'roviga to'liq mos kelmaydigan, noo'rin (irrelevant) yoki ortiqcha ma'lumotlarni taqdim etishdek jiddiy kamchiliklarga ega. Shu sababli, zamonaviy kompyuter lingvistikasi va sun'iy intellekt sohasidagi eng dolzarb muammolardan biri – bu qidiruv jarayoniga ma'lumotlarning mazmun-mohiyatini (semantikasini) chuqur tushunadigan intellektual mexanizmlarni, xususan, ontologik bilimlar bazalarini amaliyotga keng joriy etish hisoblanadi.

An'anaviy axborot qidiruv tizimlarining tarixiy evolyutsiyasiga kengroq nazar tashlaydigan bo'lsak, ularning barcha rivojlanish bosqichlari asosan ma'lumotlarning sintaktik darajadagi tahliliga qurilganini ko'rishimiz mumkin. Dastlabki davrlarda yaratilgan va ko'p yillar davomida sanoat standarti bo'lib kelgan Boolean model qidiruv so'rovlarini AND, OR, NOT kabi mantiqiy operatorlar yordamida birlashtirishga asoslangan edi. Garchi bu usul o'zining qat'iy mantiqiy qoidalari va oson dasturlanishi bilan ajralib tursada, u hujjatlarning ahamiyatlilik darajasini aniq raqamlar orqali reytinglash, ya'ni qaysi hujjat foydalanuvchiga ko'proq kerakligini aniqlash imkoniyatiga ega emas edi. Bu usul qidiruv so'zlarini faqat “bor” (True) yoki “yo'q” (False) tamoyili asosida qidirardi. Keyinchalik, axborot qidiruvini takomillashtirish, tizimga intellektuallik qo'shish maqsadida Jerald Salton tomonidan Vektor fazo modeli (Vector Space Model – VSM) va boshqa olimlar tomonidan Ehtimollar modellari (Probabilistic models) taklif etildi [19:120-122]. Vektor fazo modelida har bir hujjat matni va foydalanuvchi kiritgan qidiruv so'rovi ko'p o'lchovli geometrik fazodagi vektorlar sifatida tasvirlanadi va ularning o'xshashligi kosinus o'lchovi kabi sof matematik funksiyalar orqali hisoblanadi. Bu usul TF-IDF (Term Frequency – Inverse Document Frequency) vaznlashtirish algoritmiga tayangan holda hujjatlarning reytingini tuzishga, muhim



soʻzlarni ajratib olishga imkon berdi va axborot qidiruvi sohasida inqilobiy oʻzgarish yasadi [14:117-120]. Shunga qaramay, bu statistik va ehtimolli modellar ham matnning faqatgina sirtqi, yuzaki – leksik qatlamiga eʼtibor qaratadi. Yaʼni, ular matnda ishtirok etgan soʻzlarning chastotasini ehtiyotkorlik bilan hisoblaydi, ammo bu belgilar ketma-ketligi orqasida qanday tushuncha, qanday asl maʼno yashiringanini umuman tushunmaydi [10:85-88]. Natijada axborot qidiruvida asrlar davomida tillarda shakllangan ikkita yirik va murakkab muammo – sinonimiya (bir xil maʼnoli, ammo turli soʻzlar) va polisemiya yoki omonimiya (bir xil yoziluvchi, ammo turli maʼnolarni anglatuvchi soʻzlar) muammolari texnologik jihatdan toʻliq hal etilmay qoladi.

Sinonimiya muammosining axborot qidiruvidagi salbiy taʼsirini yaqqol koʻrib chiqadigan boʻlsak, faraz qilaylik, foydalanuvchi tizimga “avtomobil ehtiyot qismlari” deb qidiruv soʻrovi berdi. Vektor fazo modeli qidiruv bazasini aylanib chiqib, ichida aynan shu soʻzlar qatnashgan matnlarni beradi. Biroq, tizim matnida “mashina detallari” yoki “transport vositalari komponentlari” deb yozilgan, aslyatda foydalanuvchi qidirayotgan axborot bilan aynan bir xil maʼnoni ifodalovchi muhim hujjatlarni topib bera olmaydi [8:210-215]. Chunki matematik vektor modellarida “avtomobil” va “mashina” soʻzlari kompyuter xotirasida oʻlchami va shakli turlicha boʻlgan ikkita mutlaqo mustaqil belgilar ketma-ketligini va bir-biriga aslo bogʻliq boʻlmagan mustaqil vektor oʻqlarini tashkil etadi. Oqibatda qidiruvning toʻliqlik darajasi pastlab ketadi – tizim kerakli maʼlumotlarni qamrab ololmaydi. Polisemiya va omonimiya muammosida esa voqealar rivoji mutlaqo aksincha tus oladi, yaʼni foydalanuvchi “oy” soʻzini fazoviy va astronomik obyekt, yaʼni yerning tabiiy yoʻldoshi maʼnosida qidirtirganida, tizim kalit soʻz turgan joyini koʻr-koʻrona olib kelgani sababli, yilning bir qismi yanvar, fevral kabi kalendar oylari maʼnosida kelgan barcha hujjatlarni ham aralashtirib natijalar roʻyxatiga qoʻshib yuboradi. Bu holat axborot qidiruv tizimi natijalarining aniqlik



ko'rsatkichini keskin pasaytirib yuboradi, foydalanuvchini axlat axborotlar ichida qoldiradi. Ushbu muammolarni bartaraf etish, axborot texnologiyalarida kompyuterga matnlarni shunchaki harflar yoki belgilar yig'indisi sifatida emas, balki inson tafakkuridagi kabi ma'noli obyektlar, tushunchalar tizimi sifatida munosabatda bo'lishga o'rgatish zaruratini keltirib chiqardi. Aynan mana shu ulkan lingvo-texnologik ehtiyoj butunjahon o'rgimchak to'rida Semantik Web g'oyasining tug'ilishiga va uning fundamental arxitekturasini tashkil etuvchi ontologiyalar konseptining kompyuter fanlarida mustahkam o'rin egallashiga olib keldi [11:199-200].

Axborot qidiruv jarayonida ontologik bilimlar bazasini joriy etish an'anaviy leksik tahlildan semantik-kognitiv tahlilga o'tishni anglatadi [7:37-38; 20:162-165]. Ta'kidlash joizki, zamonaviy intellektual qidiruv tizimlari faqatgina bitta tor soha bilan cheklanib qolmasligi, balki foydalanuvchilarning barcha sohalardagi so'rovlariga moslasha oluvchi universal arxitekturaga ega bo'lishi talab etiladi. Shu maqsadda, tadqiqotimiz doirasida bazaviy yoki yuqori bosqich ontologiyasi modeliga asoslangan, o'zida turli sohalarni qamrab oluvchi yaxlit semantik qidiruv arxitekturasi taklif etilmoqda.

Ushbu arxitekturada W3C (World Wide Web Consortium) ning OWL (Web Ontology Language) va RDF (Resource Description Framework) standartlaridan foydalanilib, barcha sohalar uchun umumiy bo'lgan “Vaqt”, “Makon”, “Shaxs”, “Jarayon”, “Hujjat” kabi tayanch meta-klasslar yaratiladi. Har bir aniq soha esa ushbu yuqori ontologiyaning tarmog'i sifatida o'zining ixtisoslashgan tushunchalari bilan bazaga ierarxik ravishda ulanadi. Natijada, qidiruv tizimi bir vaqtning o'zida turli soha atamalarini tushuna oladigan universal intellektual yadroga (Semantic Core Engine) ega bo'ladi.

Garchi tizim barcha sohalar axborotini qayta ishlash uchun mo'ljallangan bo'lsa-da, uning ishlash prinsipini va qidiruv samaradorligini eng murakkab



terminologik apparatga ega bo'lgan yo'nalishlar – xususan, tibbiyot sohasi ontologiyasi ustida olib borilayotgan chuqur ilmiy izlanishlarimiz misolida yaqqol ko'rsatib berish mumkin. Tibbiyot ontologiyasi ko'p tarmoqli axborot tizimida o'ziga xos sinov maydoni vazifasini o'taydi [18:12-15].

Tizimda soha, yo'nalish, terminlar va ontologik kabi sinflar hamda ularni bog'lovchi ontologik baza kabi obyekt xususiyatlari (Object Properties) mantiqiy qoliplanadi [6:8-12]. Agar foydalanuvchi tizimga terminologik so'rov bersa, tizim shu termenga taluqli bo'lgan sinonim, antonym, omonim, giponing, giperonim, meronim, xolonim ma'lumotlarini, undan tashqari shu terming tegishli bo'lgan soha va subsohalari xaqidagi ma'lumotlarni chiqarib beradi. Bu algoritmik mantiq boshqa barcha sohalar uchun ham aynan bir xil andozada ishlaydi. Qidiruv tizimi qaysi sohaga oid ma'lumot izlanayotganini ontologiya daraxti orqali avtomatik tushunadi, ya'ni tibbiyotdagi “operatsiya” bilan bank sohasidagi “operatsiya”ni bir-biridan xatosiz ajratib, noo'rin ma'lumotlar taqdim etilishining oldini oladi [12:4-6].

Ko'p tarmoqli ontologik bazaning aniq ishlashi to'g'ridan-to'g'ri O'zbek tilining tabiiy tilni qayta ishlash (NLP) texnologiyalari va turli soha korpuslarining tahlil sifati bilan uzviy bog'liqdir [4:231-240]. O'zbek tili aglutinativ til bo'lganligi sababli, uning so'z yasash va shakl o'zgartirish imkoniyatlari o'ta keng bo'lib, an'anaviy stemmerlar so'z o'zagini topishda doimiy xatolarga yo'l qo'yadi [1:12-15; 3:68]. Shu sababli, universal axborot qidiruv tizimi axborotni ontologiyaga uzatishdan oldin matnlarni quyidagi qat'iy NLP bosqichlaridan o'tkazadi:

1. **Tokenizatsiya va Lemmatizatsiya:** Barcha soha korpuslaridagi matnlar (qonun hujjatlari, tibbiy qo'llanmalar, texnik adabiyotlar) so'zlarga ajratilib, har bir so'zning sof lug'aviy o'zagi (lemmasi) ajratib olinadi [5:1790-1793; 13:82-85].

2. **Morfologik va sintaktik tahlil (POS-tagging and Parsing):** So'zlarning grammatik xususiyatlari, gapdagi o'rni va tobeklik munosabatlari



aniqlanadi [2:1-4]. Bu jarayon terminlarning soha doirasidagi o'ziga xosligini aniqlashda juda muhim.

3. **Nomlangan obyektlarni tanish (NER) va ontologiyaga bog'lash (Entity Linking):** Matndan shifoxona nomlari, dori turlari, moliyaviy muassasalar yoki ilmiy terminlar avtomatik ajratib olinib, universal ontologiyadagi tegishli soha sinflariga to'g'ridan-to'g'ri biriktiriladi [15:2-5; 17:110413].

XULOSA. Olib borilgan ilmiy-tadqiqot ishlari va kompyuter tajribalari natijalariga tayangan holda shuni qat'iy ta'kidlash mumkinki, axborot qidiruv jarayonida ko'p tarmoqli ontologik bilimlar bazasidan foydalanish an'anaviy tizimlarga nisbatan mislsiz darajada yuqori samaradorlikni ta'minlaydi [14:405]. An'anaviy statistik va leksik modellar matnni faqat sirtqi belgilari bo'yicha izlash bilan chegaralanib, o'zbek tili kabi murakkab agglutinativ tillardagi sinonimiya va omonimiya muammolari oldida ojiz qolayotgan bir paytda, ontologiyaga asoslangan semantik qidiruv axborotga kognitiv yondashuvni taklif etadi. Tizim qidiruv so'rovini ko'r-ko'rona kompyuter xotirasidan izlamaydi, balki foydalanuvchining asl muddaosini mantiqiy xulosa chiqarish orqali tushunib, bilimlar grafi vositasida eng to'g'ri ma'lumotlarni topib beradi.

Tadqiqotimiz doirasida ishlab chiqilgan barcha sohalarni qamrab oluvchi universal ontologiya arxitekturasi turli domenlar orasidagi semantik aloqalarni bitta yagona intellektual yadroga birlashtira olishini amalda isbotladi [9:45-50].

Ushbu tadqiqot nafaqat nazariy kompyuter lingvistikasi uchun, balki O'zbekiston Respublikasida jadal rivojlanayotgan raqamli iqtisodiyot platformalari, elektron hukumat tizimlari va elektron ta'lim resurslari faoliyatini mukammallashtirish uchun muhim amaliy ahamiyatga ega. Kelgusida mazkur universal ontologiya bazasini yanada kengaytirish, unga yangi soha terminologiyalarini avtomatlashtirilgan tarzda o'rgatish hamda uni davlat

miqyosidagi yirik axborot portallarining asosiy qidiruv dvigateliga aylantirish bizning istiqboldagi eng asosiy ilmiy va amaliy vazifalarimizdan biri bo‘lib qoladi.

Foydalanilgan adabiyotlar ro‘yxati

1. Abdurakhmonova, N. (2016). The bases of automatic morphological analysis for machine translation. *Izvestiya Kyrgyzskogo gosudarstvennogo tekhnicheskogo universiteta*, 2(38), 12-7.
2. Abdurakhmonova, N. (2019). Dependency parsing based on Uzbek Corpus. In *Proceedings of the International Conference on Language Technologies for All (LT4All)*.
3. Abdurakhmonova, N., & Tuliyeu, U. (2018). Morphological analysis by finite state transducer for Uzbek-English machine translation. *Foreign Philology: Language, Literature, Education*, 3, 68.
4. Abdurakhmonova, N., Tuliyeu, U., Ismailov, A., & Abduvahobo, G. (2022). Uzbek electronic corpus as a tool for linguistic analysis. In *Компьютерная обработка тюркских языков. Turklang 2022* (pp. 231-240).
5. Abdurakhmonova, N. Z., Ismailov, A. S., & Mengliev, D. (2022). Developing NLP tool for linguistic analysis of turkic languages. In *2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)* (pp. 1790-1793). IEEE.
6. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021). Uzwordnet: A lexical-semantic database for the uzbek language. In *Proceedings of the 11th Global Wordnet conference* (pp. 8-19).
7. Aripov, M., Sharipbay, A., Abdurakhmonova, N., & Razakhova, B. (2018). Ontology of grammar rules as example of noun of Uzbek and Kazakh languages. In *Abstract of the VI international conference “Modern problems of applied mathematics and information technology al-Khorezmiy”* (pp. 37-38).



8. Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology behind Search* (2nd ed.). Addison-Wesley Professional.
9. Buitelaar, P., Cimiano, P., & Magnini, B. (2005). *Ontology learning from text: methods, evaluation and applications*. *Frontiers in Artificial Intelligence and Applications*, 123, 3-12.
10. Croft, W. B., Metzler, D., & Strohman, T. (2010). *Search engines: Information retrieval in practice*. Addison-Wesley Reading.
11. Gruber, T. R. (1993). *A translation approach to portable ontology specifications*. *Knowledge Acquisition*, 5(2), 199-220.
12. Guarino, N. (1998). *Formal ontology and information systems*. In *Proceedings of FOIS* (Vol. 98, pp. 3-15).
13. Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed. draft). Pearson.
14. Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.
15. Mengliev, D., Abdurakhmonova, N., Allamov, O., Ibragimov, B., Saidov, B., & Boltayev, N. (2025). *Development of a hybrid algorithm for identifying named entities in 20th century Uzbek texts*. In *AIP Conference Proceedings* (Vol. 3356, No. 1, p. 050002). AIP Publishing LLC.
16. Mengliev, D., Barakhnin, V., & Abdurakhmonova, N. (2021). *Development of intellectual web system for morph analyzing of uzbek words*. *Applied Sciences*, 11(19), 9117.
17. Mengliev, D., Barakhnin, V., Abdurakhmonova, N., & Eshkulov, M. (2024). *Developing named entity recognition algorithms for Uzbek: Dataset insights and implementation*. *Data in Brief*, 54, 110413.



18. Noy, N. F., & McGuinness, D. L. (2001). Ontology development 101: A guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, 1-25.
19. Salton, G., & McGill, M. J. (1983). Introduction to Modern Information Retrieval. McGraw-Hill.
20. Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: principles and methods. Data & knowledge engineering, 25(1-2), 161-197.