



*Gulruksor XOLIYOROVA,*  
*O'zbekiston Milliy universiteti o'qituvchisi*

*DSc, professor S.Zokirova taqrizi asosida*

### MODELING LINGUISTIC DATABASE OF UZBEK THESEURUS

Аннотация

This article scientifically and analytically examines the problems of artificial intelligence in determining the meanings of words and explaining the meanings in related forms and their solutions. The article is based on a scientific comparison of important issues such as the role of WordNet and thesaurus, thematic connections in them, the importance of thesaurus in information retrieval systems, conducted by researchers. Researchers have mainly studied how WordNet and thesaurus can help in the definition of meaning in semantic concepts and related forms, as well as how they can be used in artificial intelligence technologies. In addition, it deals with the methods, standards, and principles of their implementation used in the process of creating the thesaurus of the Uzbek language, and illuminates how artificial intelligence is developing in understanding and forming the concepts of word meanings and their related forms.

**Key words:** thesaurus, Word Net, conceptual distance, ideographic dictionary, monolingual thesaurus

### МОДЕЛИРОВАНИЕ ЛИНГВИСТИЧЕСКОЙ БАЗЫ ДАННЫХ УЗБЕКСКОГО ТЕЗАУРУСА

Аннотация

В данной статье научно-аналитически рассматриваются проблемы искусственного интеллекта при определении значений слов и объяснении значений в родственных формах и их решения. Статья основана на научном сравнении важных вопросов, таких как роль WordNet и тезауруса, тематические связи в них, значение тезауруса в информационно-поисковых системах, проведенное исследователями. Исследователи в основном изучали, как WordNet и тезаурус могут помочь в определении значения семантических понятий и связанных с ними форм, а также как их можно использовать в технологиях искусственного интеллекта. Кроме того, рассмотрены методы, стандарты и принципы их реализации, используемые в процессе создания тезауруса узбекского языка, и освещено, как развивается искусственный интеллект в понимании и формировании понятий о значениях слов и родственных им формах.

**Ключевые слова:** тезаурус, WordNet, концептуальная дистанция, идеографический словарь, одноязычный тезаурус.

### O'ZBEK TILI ELEKTRON TEZAVURUSINI MODELLASHTIRISHNING LEKSIKOGRAFIK TADQIQI

Аннотация

Mazkur maqolada sun'iy intellektning so'z ma'nolarini aniqlash va ma'nolarni bog'liq shakllarda tushuntirishdagi muammolar va ularni hal qilish usullarini ilmiy-analitik ravishda ko'rib chiqadi. Maqola tadqiqotchilar tomonidan olib borilgan WordNet va tezavurusning o'rni, ulardagi mavzui bog'lanishlar, ma'lumotlarni izlash tizimlarida tezavurusning ahamiyati kabi muhim masalalarni ilmiy jihatdan taqqoslashga asoslangan. Tadqiqotchilar asosan WordNet va tezavurusning ma'nolar tushunchalari va bog'liq shakllardagi ma'noni aniqlashda qanday ko'maklashish mumkinligini, shuningdek, sun'iy intellekt texnologiyalarida ularni qanday qo'llash imkoniyatlarini o'rganishganlar. Bundan tashqari o'zbek tilining tezavurusini yaratish jarayonida qo'llaniladigan metodlarga, standartlarga, va ularning amal qilish prinsiplariga duch keladi hamda sun'iy intellektning so'z ma'nolari va ularning bog'liq shakllardagi tushunchalarini tushunish va shakllantirishda qanday rivojlanayotganligini yoritadi.

**Kalit so'zlar:** tezaurus, Word Net, konseptual masofa, ideografik lug'at, monolingval tezaurus

**Kirish.** Ma'lumki, tezaurus borasida dunyo tajribasida ko'plab izlanishlar amalga oshirilgan. Tezaurus bugungi kunda, ayniqsa, kompyuter texnologiyalari va sun'iy intellekt rivojlangan paytda tezaurus terminlar asosida birlashgan konseptlar asosidagi nazoratga olingan va strukturalashgan lug'at hisoblanadi.

Ayniqsa, umumiy tilshunoslik nuqtai nazaridan tezaurus terminiga doir tushuncha kognitiv ahamiyat kasb etib, S.Osokinaning ta'biricha, tilshunoslik tizimiga oid boshqa tushunchalar, ya'ni "kontseptual tizim", "dunyo tasviri", "aqliy leksika" kabilarda o'z ifodasini topadi. Y.N.Karaulov asarlarida tezaurus tushunchasining evolyutsiyasi, tezaurus yondashuvining rivojlanishi madaniyatshunoslikda psixologiya, sotsiologiya va adabiyotshunoslik fanlari misolida ko'rib chiqiladi. Ushbu fan sohalarida tezaurusni kognitiv paradigmasiga mos ravishda izohlaydi, uni odamlarga xos bo'lgan bilimlarning aqliy tizimi bilan aniqlaydi. Tezavurusning til tomonidan tashkil etilgan bilimlar tizimi sifatida taklif etilayotgan kontseptsiyasida prinsipial yangilik shundan iboratki, tezavurusning lingvistik (aqliy emas) tabiati o'rganiladi, shuningdek, ba'zi psixik tuzilmalarning gnoseologik ikkilamchi tabiati bilan bog'liqligi o'rganiladi.

George A. Miller va boshqalar tomonidan WordNet yaratishga doir olib borgan kuzatishlariga ko'ra, 1928-yilda yaratilgan "Murray's Oxford English Dictionary" lug'ati diaxronik tamoyillarga asoslanib, undagi so'zlarning berilishi ham

tarixiy aspektga tayanilgan. Olimlar tomonidan psixoleksikologiya termini ishlatiladi. 20-asrda psixolingvistik fan olamiga kirib kelishi natijasida lingvistik kompetensiyaning kognitiv asoslari bilan integrallashuvi sodir bo'lganligini yuqorida nomlari zikr etilgan olimlar tomonidan qayd etiladi. Miller va Jonson Lairdning (1976) leksik bilimlarni hamda zamonaviy lingvistik strukturani aniqlashda jiddiy faktorlarlarni inobatga olgan holda psixoleksikologiya deb nomlanishi lozimligini e'tirof etishadi. Shu bilan birga leksikon uchun beriladigan ma'lumot fonologiya, sintaksis hamda leksik komponentlardan iborat bo'lib, xabarlardagi kunlik hosilalar bilan birga ishlanishi lozimligini bu esa o'z navbatida psixolingvistika bilan hamkorlik qilish lozimligini taqozo etadi. Bu shuni anglatar ediki, lug'at leksikoni uchun kiritiladi leksik birliklar assotsiatsiya bilan uyg'unlashishi leksikografiyada mental leksikonni ko'plab sinxron belgilarini kashf etish lozim.

Rus tilshunosligida S.Osokina lingvistik tezaurus yaratishga asoslari bo'yicha tadqiqot olib borgan. Shuningdek, I. B. Azarova, H. B. Lukashевич, N. Gader, M. Benjamin, M. Rosman kabi olimlar ham tezavurusning turli aspektlarini tahlilga tortgan. Bu borada S.Osokina tezaurus lug'atlarni quyidagicha tasniflaydi: so'z bilan so'zning ma'no munosabatlarini ochib beruvchi lug'at (sinonim, antonym so'z birikmasi); 2) tushunchadan so'z bilan bog'liq munosabatni ochib beruvchi

lug'at; (Roje tezaurusi, O.S.Baranovning Ideografik lug'ati, WordNet, RussNet, RuThes electron tezauruslar); 3) stimul so'zlardan reaksiya bildiruvchi so'zlar munosabatini ochib beruvchi lug'at (Assotsiativ lug'at). Yuqoridagilardan kelib chiqib olina tezaurusga quyidagicha ta'rif beradi:

- 1) uyadosh so'zlar asosidagi leksik tizim;
- 2) mavjud bilimlardan ifodalovchi tushunchalar tizimi;
- 3) psixolingvistik tajriba usuli bilan qurollanuvchi hamda inson tafakkurida mavjud bilimlar tizimi.

Darhaqiqat, tezaurus ideografik lug'at sifatida inson idroki orqali anglashiladigan tasavvurlar va ularning lisoniy modeli orqali namoyon bo'luvchi ma'lumotlar bazasi hisoblanadi. Shuning ta'kidlash o'rinliki, tezaurus insoniyat tomonidan bir necha asrlar davomida shakllangan assotsiativ verbal vositalar orqali anglashilgan konseptlar yig'indisi hamdir. Manvalarda qayd etilishicha, inson bolalik davridan boshlab oilaviy muhit va ijtimoiy ta'sir orqali atrof muhit haqidagi tasavvurlarni o'zining lisoniy xotirasida gavalantiradi. Maktab va maktabdan keyingi ta'limda bolaning tasavvuri ilmiy ma'lumotlar bilan boydii. Ijtimoiy tushunchalar bilan birga uning akademik bilimlari ham ixtisoslashib boradi. Mehnat faoliyati, qolaversa, kishilik jamiyati bilan yanada turli munosabatlar, ya'ni qarindoshlik, xodim, qolaversa, muayyan mas'uliyat asosidagi erishgan ma'lumotlari asosida unda yanada chuqurroq tasavvur lisoniy model orqali hosil bo'ladi. Bu esa uning hayoti davomida to'plagan tezaurusni yaratishga yordam qiladi. Muayyan tushunchalar qaysidir tushunchalar ostida birlashadi. Xotira va psixolingvistik jarayon orqali erishilgan barcha ma'lumotlar tezaurusning mohiyat nuqtai nazaridan asosini tashkil qilishda omil bo'lib hizmat qiladi.

**Metodologiya.** Endilikda barcha rivojlangan tillarda tezauruslarning elektron shakllari yaratilgan va ularning imkoniyatlari tabora kengayib bormoqda.

RussNet va RuThes rus tili bo'yicha yaratilgan tezauruslar bo'lib, ularning ayrimlarini alohida semantik tasnifiy xoslanishini ko'rib o'tamiz.

Rus tilining RussNet tezaurusi loyihasi 1999-yilda Sankt Peterburg universiteti Filologiya fakulteti Amaliy va matematik lingvistik kafedrasida tomonidan amalga oshirila boshlangan. Ushbu lug'at leksik semantik lug'at sanalib, uning lingvistik imkoniyatlariga rus tilining barcha leksik qatlamini kiritish, zamonaviy rus tilining asosiy lug'at bazasini qamrab olish, shuningdek, rus tilidagi so'zlar o'rtasidagi semantic, semantik-grammatik, semantik-derivatsion munosabatlarni ifodalash asosiy maqsad qilib olingan.

Ushbu lug'at WordNet shablonidan namuna sifatida olinib, to'rt so'z turkumdan iborat: ot, fe'l, sifat, ravish. Rus tilidagi leksikalashgan konsept asosida sinonimik qatorlar (sinset)dan iborat. Sinsetlar paradigmatic va sintagmatic munosabatlardan iborat. Uning hajmi 5500 sinset bo'lib, 1300 ot sinset, 1900 fe'lga oid sinset, 1100 ta sifatga oid sinset hamda 200 ta ravish sinsetdan iborat. RussNetdagi munosabatlar ikki guruhga ajratilgan:

sinsetlar o'rtasidagi munosabatlar  
so'zlar o'rtasidagi munosabatlar

Lug'atdagi sinsetlar o'z navbatida quyidagi munosabatlarga ko'ra farqlangan:

- 1) giponim va giperonimlar;
- 2) antonimlar;
- 3) konversiyalar;
- 4) meronim yoki holonimlar;
- 5) entailment:
- a) presuppozitsiya;
- b) subevent;
- c) kauzatsiya;
- d) inchoative/terminative (boshlanish yoki tugashi).
- 6) munosabatga jalb qilingalik yoki munosabatdagi roli;
- 7) sinonimiyaga yaqin so'zlar.

So'zlar o'rtasidagi munosabatlarga quyidagilar kirishi qayd etilgan:

- 1) sinonimiya;
- 2) derivatsion sinonimiya;
- 3) derivatsion giponimiya;

4) derivatsion munosabatlarga kirishganlik va uning o'rni;

5) aspect\_oppozitsiya

Demak, yuqorida sanab o'tilgan turlar bu tezaurus ichidagi munosabatlar sanaladi.

Ingliz tili manbalarda tezaurus lug'atlarning quyidagi turlari ko'rsatiladi:

Axborot qidiruv tezauruslari axborot qidiruvi tizimi uchun yo'naltirilgan bo'ladi. Unda xoh intellektual xoh avtomatik ravishda bo'lsin ma'lumotlar azasini indekslashda yoki uni qidirishda tezaurusdan foydalaniladi. Bu esa to'rt imkoniyatni taqdim qiladi:

- Tezaurus ham ma'lumot qidirishda ham indekslashda foydalaniladi;

- Tezaurus qidirishda emas faqat indekslashda foydalaniladi;

- Tezaurus indekslashda emas qidiruv uchun qo'llanadi;

- Tezaurus har ikkisida ham ishtirok etmaydi.

Kitob mualliflari tezaurusning ham qidiruv, ham indekslashdagi imkoniyatga ega lug'atni klassik tezaurus deb nomlaydi. Bunda tezaurusdagi faqat stemlar, ya'ni o'zak so'zlar asosiy vazifa bajaradi. Bu kabi tezauruslar indeks uchun qo'llanganda ma'lumotlar bazasi kompilyatorlari orqali amalga oshiriladi, axborot qidiruvda esa tezausuru ma'lumot qidiruv instrumenti vazifasini bajaradi. An'anaviy tezauruslar internet axborot tizimida qo'llanadi. Mualliflar bunda ko'plab media tarmoqlaridan tezaurus uchun zarur ma'lumotni o'zining bazasiga o'tkazib olinishini, qolaversa, taksonomiyalar bilan birgalikda lingvistik inferensi inferens sifatida (asosan statistic va qoidaga ko'ra) dasturiy ta'minotning yangi generatsiyasi hosil qilinishini ta'kidlaydi. Bu borada mualliflar ushbu fikrlari asosli bo'lib, taksonomiya tezaurusdagi bilimlar bazasini tizimli tahlil qilish imkoniyatini yaratadi. Ayniqsa, ushbu manbada bildirilgan ayrim mulohazalar o'rinni keltirilgan. Unga ko'ra ma'lumotlar bazasidagi matnli hujjatlarni tahlil qilish asosida sohaviy kategoriyalar avtomatik ravishda generatsiya qilish imkonini yaratadi. Garchi statistic ravishda ma'lumotlarni to'plash va ularni tezaurus sifatida qidirish imkoniga ega bo'lsa-da, natijalarni fine tuning qilish hamda tahrirlash uchun dasturiy ta'minot uchun inson ishtiroki zarur bo'ladi. Natijada tezaurus va klassifikatsiyani tuzishda joriy tijorat maqsadga yo'naltirilgan dasturiy vositalarni kompyuter yordami sifatida qarash mumkin bo'ladi.

Mualliflar tomonidan keltirilgan indekslovchi tezauruslar qidiruv uchun emas, balki indekslashga mo'ljallangan bo'ladi.

Leksikografik tadqiqotlarda tezaurus va WordNet yaratishga doir ko'plab tadqiqotlar amalga oshirilgan. Manbalarda qayd etilishicha, WordNet boshqa lug'atlardan farqli ravishda besh kategoriyaga ko'ra bo'linadi. Odatda an'anaviy lug'atlarda ot, fe'l, sifat, ravish va funksional so'zlar (yordamchi so'zlar) mavjud bo'lsa WordNetda faqat shulardan to'rttasi, ya'ni ot, fe'l, sifat va ravish so'z turkumiga tegishli birliklar kiritiladi. WordNetga kiritiladigan so'zlar maxsus sinsetlarga birlashtirilib, yordamchi so'z turkumiga tegishli birliklar tilning sintaktik komponenti sifatida qaralgan.

Wordnetda ot so'z turkumiga kiruvchi birliklar leksik xotirada mavzuiy iyeraxiya sifatida asos sanaladi. Fe'l so'rkumidagi birliklar otga ergashuvchi sifatida, sifat va ravish esa har ikkisining belgini igodalovchi parameter sanaladi. Har bir leksik struktura turli tasnifiy hodisani ifodalanishi, barcha sintaktik kategoriyalar uchun har bir alohida tashkil qilingan tamoyil leksik bilimlarni psixologik murakkablashtishda noto'g'ri ifodalanishi Miller va boshqal tomonidan olib borilgan kuzatishlarda qayd etiladi. WordNetning o'ziga xos jihati shundaki, sozlarning so'z shakliga ko'ra emas, balki ma'nosiga qarab leksik ma'lumotlarni loyihalashtirishga harakat qilinadi. Shu jihatdan WordNet an'anaviy lug'atlardan ham ko'ra tezaurusga o'xshaydi.

O'zbek tili tezaurusiga doir tadqiqotlar sirasiga ilk bor UzWordNet ga yaratishga muvaffaq bo'lgan olimlarning tadqiqotlariga murojaat etmoqchimiz.

Ushbu lug'at Prinston lug'ati asosida yaratilib, 20683 so'zning 64389 ma'nosi hamda ularning 28140 sinsetdan iborat bazasi kiritilgan. Ushbu lug'at ingliz tili asosida qilingan bo'lib, so'zlarning aniqlik darajasi 75.98%ni tashkil qiladi. Ushbu olimlar olib brogan tadqiqot uch bosqichni qamrab olgan:

- Leksik resursni qayta ishlash;
- O'zbek tili adabiy tili uchun Prinston WordNeti asosida tezaurus strukturasi avtomatik yaratish;
- Avtomatik qurilgan lug'atni inson yordamida tekshirish va baholash

Ushbu tadqiqotning 1-bosqichida tarjima lug'at asosida so'zlarning tarjima bazasi yaratilgan. Ingliz-o'zbek qismi uchun 40000 so'z, o'zbek-ingliz qismida 30000 lemma mavjud. Ushbu

tarjima lug'at kitob shaklida bo'lgani bois uning eltron shaklga o'tkazishda quyidagilar bajarilgan:

- Lug'at mashina formatiga o'tkazilgan;
- Jadval ko'rinishdagi lug'at uch ustundan iborat bo'ldi:
  - a)ingliz tilidagi lemma;
  - b)lemmaning so'z turkumi;
  - c)o'zbek tilidagi tarjimasi

Xulosa o'rnida shuni e'tirof o'rinliki, o'zbek tili tezaurusi uchun yaratiladigan ma'lumotlar bazasi tabiiy tilni qayta ishlash jarayoni uchun, qolaversa, axborot qidiruv tizimi uchun moslanganligi kelgusida sun'iy intellekt texnologiyalarida tadqiqotlar uchun muhim manba bo'lib xizmat qiladi.

#### ADABIYOTLAR

1. Ожегов С.И., Шведова Н.Ю. Толковый словарь русского языка. – Москва, 2006. – С. 944.
2. Розентал Д.Э., Теленкова М.А. Словар-справочник лингвистических терминов (Пособие для учителя). – Москва: Просвещение, 1985. – 399 с.
3. Петровский Н.А. Словарь русских личных имен. Изд. 4-е, дополн. - М.: Русские словари, 1980. –С. 414
4. Ўзбек тилининг изоҳли луғати. 2-жилд. – Тошкент: Ўзбекистон миллий энциклопедияси, 2006. – Б.492.
5. Ўзбек тилининг изоҳли луғати. 5 жилдлик. 4-жилд. –Тошкент: ЎзМЭ, 2006. – 671 б.
6. Ҳоҗиев А. Лингвистик терминларнинг изоҳли луғати. – Тошкент, 1985. – 67-б.
7. Matlatipov San'atbek, Aripov Mirsaid, Abdurakhmonova Nilufar. (2018).
8. Modeling WordNet Type Thesaurus for Uzbek Language Semantic Dictionary. International Journal of Systems Engineering, 2(1), 26-28.
9. <https://doi.org/10.11648/j.ijse.20180201.16>
10. Agostini, A., Usmanov, T., Khamdamov, U., Abdurakhmonova, N., & Mamasaidov, M. (2021, January). Uzwordnet: A lexical-semantic database for the uzbek language. In Proceedings of the 11th Global Wordnet conference (pp. 8-19).
11. Nilufar, A., & Gulruhsor, H. (2024, May). O'zbek tili tezaurusini yaratishda leksik birliklarning ifodalanish. In Konferensiyalar| Conferences (Vol. 1, No. 10, pp. 186-192).
12. Nilufar, A., & Gulruhsor, X. (2024). O'zbek tili tezaurusini yaratishda standart me'yorlar. «Contemporary technologies of computational linguistics», 2(22.04), 464-469.