

## **O‘ZBEK TILI TEZAURUSINI YARATISHDA LEKSIK BIRLIKLARNING IFODALANISH**

**Abduraxmonova Nilufar,**  
O‘zbekiston milliy universiteti  
Kompyuter lingvistikasi va amaliy  
tilshunoslik kafedrasini mudiri,  
filologiya fanlari doktori, professor

**Holiyorova Gulruhsor,**  
O‘zbekiston milliy universiteti  
Kompyuter lingvistikasi va amaliy  
tilshunoslik kafedrasini  
o‘qituvchisi

**Annotatsiya:** Ushbu maqolada tezaurus yaratishning ahamiyati, tadqiqotlar, va olimlar tomonidan qo‘llangan usullar tahlilga tortilgan. Tezaurusning leksik birliklarini yaratishda, shuningdek, o‘zbek tilidagi sinonim, antonim, omonim, va taksonomik munosabatlarning yanada keng jihatlaridan foydalanishning muhim jihatlari ochib berilgan. Bundan tashqari tezaurus yaratishning bir nechta usullari va standartlari o‘zbek tili uchun muhimligi qayd etilgan. Standartlarga mos ravishda tezaurus yaratishning leksik birliklarni ifodalashdagi xususiyatlari ko‘rsatiladi. So‘nggi qismda o‘zbek tili tezaurusining qanday yaratilishi va undagi ma‘lumotlar bazasining axborot qidiruv tizimlari uchun muhim jihatlari bayon etilgan.

**Tayanch so‘zlar:** tezaurus, WordNet, konseptual masofa, ideografik lug‘at, monolingval tezaurus.

**Abstract:** This article analyzes the significance of thesaurus creation, research, and methods used by scholars. The importance of using the thesaurus in the creation of lexical units, as well as the wider aspects of synonyms, antonyms, homonyms, and taxonomic relations in the Uzbek language, is shown. In addition, the importance of several methods and standards of thesaurus creation for the Uzbek language is analyzed. Features of creating a thesaurus in the representation of lexical units are shown in accordance with the standards. In the last part, it is analyzed how the thesaurus of the Uzbek language is created and how its database can be an important resource for information search engines.

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

**Аннотация:** В статье анализируется значение создания тезауруса, исследования и методы, используемые учеными. Показана важность использования тезауруса при создании лексических единиц, а также более широкие аспекты синонимов, антонимов, омонимов и таксономических отношений в узбекском языке. Кроме того, показана важность некоторых методов и стандартов создания тезаурусов для узбекского языка анализируется. Показаны особенности создания тезауруса в представлении лексических единиц в соответствии со стандартами. В последней части анализируется, как создается тезаурус узбекского языка и как его база данных может быть важным ресурсом для информационных поисковых систем.

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

Axborot qidiruv tizimi uchun tezaurusning ahamiyati katta. Foydalanuvchi tomonidan yuboriladigan so‘rov orqali olinadigan ma’lumotlar semantik to‘rga, ya’ni tezaurusga bog‘langan bo‘ladi. So‘zlar va konseptlar o‘rtasidagi leksik va semantik munosabatlarni formal vositalar orqali ifodalash mumkin [Edward A.F. , Nutter J.T. etc. Building a large thesaurus for information retrieval // - P. 101 ]. Lingvistik lug‘atlarda so‘zlarning leksik sematik jihatdan munosabatlari sinonim, antonim, omonim kabi munosabatlari bilan cheklanib qoladi, biroq axborot qidiruvi uchun so‘zlarning taksonomik va iyerarhik munosabatlarining ham berilishi katta ahamiyatga ega. J. Edward va boshqalar tomonidan olib borilgan tadqiqotlarda axborot qidiruv tizimi uchun katta hajmdagi tezaurus yaratishda mashina o‘qiy oladigan lug‘atlar uchun matnlarni qayta ishlash metodidan foydalanib, maqolalarga berilgan izohlar analiz qilingan. Semantik tarmoqdagi munosabatlar bilan olingan ma’lumotlar birlashtirilgan. So‘ng Smart va CODER eksperimental qayta ishlash tizimidan foydalanib, tezaurusdagi birliklar tekshirilgan.

J.Aitchison, A.Gilchrist va D.Bawden kabi olimlar tomonidan tezaurus modellari bo‘yicha izlanishlar olib borilgan. Ularning “Thesaurus construction and use: a practical manual” [J.Aitchison, A.Gilchrist va D.Bawden Thesaurus construction and use: a practical manual (4-edition), Taylor & Francis e-Library, 2005. -202 p.] nomli qo‘llanmasida tezaurusga doir ilmiy tadqiqotlar tahlilga tortilgan. Unga ko‘ra axborot qidiruv tizimida tezaurus 4 xil imkoniyatga ega bo‘lishi mumkin:

- tezaurus qidiruv va indeksatsiya uchun foydalaniladi;
- tezaurus indeksatsiyada foydalaniladi, qidiruvda foydalanilmaydi;
- tezaurus qidiruvda foydalaniladi, indeksatsiyada foydalaniladi;
- tezaurus har ikkisida ham foydalanilmaydi.

Tezaurus yaratishning bir nechta usullari mavjud. Yuqorida tadqiqotchilar tomonidan keltirilgan tezaurus turlariga qarab ularning qurilish arxitekturasi va kontenti ham talablari ko‘p tilli yoki bir tilli tezaurus uchun standartlar yaratiladi. Dastlab 1974-yilda Xalqaro standartlashtirish tashkiloti tomonidan bir tilli tezaurus uchun yaratilgan. Keyingilari 1985-yillarda nashr qilingan. 1967-yilda nashr qilingan Injineriya va ilmiy terminlar Amerika tezaurusi alifbo tartibida tuzilgan. Monolingval tezaurus uchun milliy standartlar Britaniya standartini o‘z ichiga olgan. 1959-yildan 1993-yilgacha bo‘lgan evolyutsion yo‘riqnomasi va tezaurus tuzish tamoyillarining shakllanish bosqichlari Krooks va Lankaster tomonidan tadqiq etilgan[J.Aitchison, A.Gilchrist va D.Bawden Thesaurus construction and use: a practical manual (4-edition), Taylor & Francis e-Library, 2005. – P. 25 ].

J.Aitchison fikricha, tezaurus yaratishning dastlabki bosqichlaridan biri bu uning tarkibiga kiritiladigan birliklarning grammatik shakli, talaffuzi, ko‘plik shakli, qisqartma yoki qo‘shma shaklda ekanligi tashkil etadi. Ikkinchi bosqichda u yoki bu konseptga tegishli konseptni ifodolovchi ikki yoki undan ortiq sinonimlarning mosini tanlash masalalari o‘rganiladi. Uchinchi bosqichda muayyan turga kiruvchi terminlarni, jumladan o‘zlashma so‘zlar, slenglar, tijorat nomlar va atoqli otlarni aniqlash kabilar bajariladi. To‘rtinchi, tezaurusning xususiy maqsadidan kelib chiqqan holda terminlarning turli kontekstda u yoki bu ma’noda qo‘llanishini cheklash uchun ayrim qoidalar yaratiladi. Ushbu cheklov omonimlarni aniqlash imkonini beradi. Ingliz tili terminologiyasida *controlled language nazorat qilingan tilda* indekslangan termini tanlangan (preffered yoki non preferred term) termin sifatida qo‘llanadi. Tanlangan termin berilgan konseptni ifodalashda ishlatiladi. U deksriptor yoki kalit so‘z deb ham nomlanadi. Tanlanmagan termin sinonim yoki kvazi sinonim sifatida ham qo‘llanadi. Ushbu birliklar indekslash uchun emas, balki tanlangan terminga mos birlikni aniqlashga yordam beradi.

Mualliflarning qayd etishicha, indekslangan terminlar orqali ifodalangan konseptlar umumiy kategoriyalarga tegishli bo‘ladi. Ushbu kategoriyalar fasatlar analizida fundamental kategoriyalardan iborat bo‘ladi. Ular konkret birliklar, mavhum konseptlar va atoqli otlardan iborat bo‘ladi.

Konkret birliklar:

Narsalar va fizik qismlar:

BOSH QISM

BINOLAR

QAVATLAR

OROLLAR

Materiallar

SEMENT

YOG‘OCH

ALYUMIN

Abstrakt konseptlar

Faoliyat va voqealik

EVOLYUTSIYA

BOSHQARUV

URUSHLAR

MAROSIMLAR

Abstrakt birliklar, narsalar, materiallar, faoliyatlar

QONUN

NAZARIYA

KUCHLANISH

**SAMARADORLIK**

Sohalar va fanlar

**FIZIKA**

**METEROLOGIYA**

**PSIXOLOGIYA**

O‘lchov birliklari

**KILOGRAM**

**METR**

Atoqli otlarga tegishli Individual birliklar

**NIGERIA**

**INSON HUQUQLARI KOMISSIYASI**

Shu tarzda standart birliklarning kategoriyalari keltiriladi. Shuningdek, indekslangan terminlarga ot va otli frazalar keltiriladi. Otli frazaga kiruvchi umumiy shakllar sifat frazadan iborat bo‘ladi.

Masalan,

**AYOL ISHCHILAR**

**BOSMA NASHRLAR**

**SHAHAR JAMOAT JOYLARI**

Standartlarga ko‘ra, garchi ayrim tezauruslarda vaqt, holat, o‘lcham, hajm, vaziyat va xususiyat kabil sifatlar kiritilgan bo‘lsa-da, indekslanuvchi terminlar sirasiga sifatlar qabul qilinmagan. ANSI standartlarida sifatlarni alohida holatda ko‘rsatish tavsiya qilinmaydi.

Olimlar ushbu fikrga namuna sifatida quyidagilarni keltirishadi:

<PORTABLE=> portable typewriter

USE TYPWRITER and AUXILIARY>

RECTANGULAR=> Rectangular windows

USE WINDOWS AND RECTANGULAR>

Shuningdek, ayrim standartlarda ularni otli birikmalarning tarkibi sifatida keltirish o‘rinli deb qaralgan:

**MINIATURE SIZE**

**PORTABLE DEVICES** kabi.

Ravish so‘z turkumiga tegishli birliklar, ayniqsa *very*, *highly* kabilar tezaurusga kiritilmaydi, ular faqat muayyan birikmaning tarkibida keltiriladi: *very large scale integration*

Tezaurusda fe‘l so‘z turkumi ham kiritilmaydi, uning o‘rniga u hosil qilgan ot, fe‘l ishtirokidagi otli birikma, gerundiy kabilar kiritiladi: **COMMUNICATION** (**COMMUNICATE** emas), **ADMINISTRATION** (**ADMINISTER** emas), **WALKING** (**WALK** emas).

Ingliz tilida yaratilgan tezauruslarda artikllar ham kiritilmaydi.

Otga tegishli birliklarning son kategoriyasida berish masalasida esa tillarning lingvistik xoslanishiga ko‘ra standartlar belgilanadi. Manbada qayd etilishicha, nemis va franzus tillari uchun ko‘plik kategoriyasida otlar tezaurusda berilmaydi, biroq ingliz tilidagi ayrim otlar ko‘plikda kelishlikni taqozo etadi va shunday beriladi. Turdosh otlarning sanaladigan va sanalamaydigan turlari ham tezaurusdan o‘rin oladi. Shu bilan birga ayrim sanaladigan va sanalamaydigan otlarga ko‘plik kategoriyasini qo‘shish orqali ularning muayyan sinfini ifodalash orqali ham ot leksemalar berilishi mumkin [Russian WordNet From UML-notation to Internet/Intranet Database Implementation Valentina Balkova<sup>2</sup>, Andrey Sukhonogov<sup>1</sup>, and Sergey Yablonsky – P.31.]: STEELS, FRUITS, CEMENTS.

Abstrakt konseptlar birlikda quyidagi turlar bo‘yicha tasniflangan:

Abstrakt birliklar va fenomenlar

AVTORITET

MANTIQ

IJTIMOIIY MASOFA

Xoslik

ECCENTTICITY

EMOTSIONAL XOSLIK

E’tiqod tizimlari

SOTSIALIZM

INDUIZM

Harakat-holat

FUZIYA

KONFLIKT

Sohalar

Biokimyo

Injineriya

Etnologiya

Agar birdan ortiq sinflar sifatida abstract konseptlar ifodalasa, bu kabi sinflar ko‘plikda berilgan:

INFORMATSION BILIMLAR

BIOLOGIK SOHALAR

IDEOLOGIYALAR

Tadqiqotchilar ANSI standartlarida yagona birliklarning birlik kategoriyasida berilishini ta’kidlaydilar: BIG BEN, RAINBOW’S END

Ham birlik va ko‘plikda keluvchi otlarning berilishini quyidagicha tavsiflaydilar. Unga ko‘ra agar tezaurusda har ikkisi mavjud bo‘lsa, ular o‘rtasidagi farqni ko‘rsatishda quyidagi holatda beriladi: COATING (jarayon) va COATINGS (material). Boshqa bir standartlarda ushbu holatlarni farqlash uchun quyidagicha

berish holatlari kuzatilishini olimlar e’tirof etishadi: COATING (jarayon) COATING PROCESS.

Tezaurusda transliteratsiya jarayoni ISO standartiga mos ravishda umum qabul qilgan talaffuz me’yorlarga ko‘ra belgilar konvertizatsiya qilinadi. Qo‘shma otlar orasidagi chiziqchalar tezaurusda olib tashlanadi. Agar qo‘shma otning bir tarkibi bir harfdan iborat bo‘lib, ma’no anglatmasa, chiziqchalar saqlanib qoladi: N-TYPE STARS. Agar qo‘shma otlar tarkibidagi bir nechta konseptlar orasidagi munosabat mavjud bo‘lsa, u holda chiziqcha saqlanadi: PROTON\_NEUTRON INTERACTION.

So‘zlarni bosh harflarda berish masalasi ham standartlarda turlicha yondashiladi: ANSI standartlarida indekslanuvchi terminlarning barchasi bosh harflarida berish maqsadga muvofiq deb qaralsa, ISO 2788 va BS 5723 bosh harflarda berish kuzatilmaydi.

Tezaurusda abreviatura va akronimlar barcha uchun tushunarli va keng iste’molda qo‘llansa, indekslanuvchi birlik sifatida kiritiladi. Agar qisqartmalar ko‘p ma’no bo‘lsa u holda quyidagicha keltiriladi: BBC/BRITISH BROADCASTING CORPORATION BBC/BROWN BOVERI COMPANY

Tezaurusda neologizm, sleng va jargonlar yangi konsept sifatida muayyan soha, submadaniyat yoki ijtimoiy guruhga tegishli terminlar sifatida standart lug‘atlarda mavjud bo‘lmaydi. Shuning uchun ular tezaurusda ular descriptor sifatida qabul qilinadi. Shuningdek, tezaurusda mashhur shaxs nomlari, ilmiy nom yoki ismlar, rasman tan olingan umumtushunchaga ega joy nomlari, shaxs va institut nomlari keltiriladi, biroq shuni ta’kidlash o‘rinliki, har bir tilning mahallaiy nomlovchi atov birliklar bo‘lib, ular global tezaurusda ifodalnmasada, milliy tezaurusda berilishi mumkin.

Bundan tashqari tezaurusga kiritiladigan leksik birliklar omonim bo‘lsa, ularning yonida ma’no farqlashga xizmat qiluvchi aniqlovchisi qavs ichida beriladi: CELLS (biologiya) CELLS (elektrik) READING (joy) READING (jarayon). Mualliflar lug‘atning berilishi borasida quyidagi jadvalni namuna sifatida keltiradi [Russian WordNet From UML-notation to Internet/Intranet Database Implementation Valentina Balkova<sup>2</sup> , Andrey Sukhonogov<sup>1</sup> , and Sergey Yablonsky – P.34.]:

| Indexing language | Indexing language | Indexing language | Indexing language      |
|-------------------|-------------------|-------------------|------------------------|
| <b>A</b>          | <b>B</b>          | <b>C</b>          | <b>D</b>               |
| Food              | Food              | Food              | Food                   |
|                   | Bakery Products   | Bakery Products   | Bakery Products        |
|                   |                   | Bread             | Bread                  |
|                   |                   |                   | <i>(By form)</i>       |
|                   |                   |                   | Bread Loaves           |
|                   |                   |                   | Bread Rolls            |
|                   |                   |                   | <i>(By ingredient)</i> |
|                   |                   |                   | → Brown Bread          |
|                   |                   |                   | White Bread            |
|                   |                   |                   | Starch Reduced Bread   |

System A—Brown Bread, USE Food

System B—Brown Bread, USE Bakery Products

System C—Brown Bread, USE Bread

System D—Brown Bread

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.

#### **FOYDALANILGAN ADABIYOTLAR:**

1. Eneko Agirre<sup>1</sup>, Enrique Alfonseca<sup>2</sup>, and Oier Lopez de Lacalle Approximating Hierarchy-Based Similarity for WordNet Nominal Synsets using Topic Signatures -P. 15 / GWC 2004 Second International WordNet Conference, GWC 2004 Brno, Czech Republic, January 20–23, 2004 Proceedings (CD-ROM version)
2. Edward A.F. , Nutter J.T. etc. Building a large thesaurus for information retrieval // - P. 101
3. J.Aitchison, A.Gilchrist va D.Bawden Thesaurus construction and use: a practical manual (4-edition), Taylor & Francis e-Library, 2005. -202 p.
4. Russian WordNet From UML-notation to Internet/Intranet Database Implementation Valentina Balkova<sup>2</sup> , Andrey Sukhonogov<sup>1</sup> , and Sergey Yablonsky – P.31.
5. Russian WordNet From UML-notation to Internet/Intranet Database Implementation Valentina Balkova<sup>2</sup> , Andrey Sukhonogov<sup>1</sup> , and Sergey Yablonsky – P.34.