

ORONIMIK QURILMALARNING MILLIY KORPUSDA BERILISH MASALASI

Toirova Guli Ibragimovna

tugulijon@mail.ru

Filologiya fanlar doktori,
BuxDU professori

Annotatsiya. Maqolada XXI asr ilm-fani bu borada elektron korpuslar misolida keng tadqiq imkonini oronimlarning milliy lisoniy tabiati, ularni lug'atlashtirish hamda milliy korpusda berish masalasiga e'tibor qaratilgan. Toponimika sohasida amalga oshirilgan ilmiy tadqiqotlar, manbalarda oronimlar va oronimika tushunchasi, shuningdek, uning o'rganish obyektni belgilash masalasida turlicha fikrlarni uchratishi sababli yagona milliy korpusda ularni aks ettirish muammosi yoritilgan. Hozirda oronimlarni avtomatik aniqlashning quyidagi latent-semantik tahlil (Latent semantic analysis), parallel matnlarni tenglashtirish (Bitext word alignment) hamda o'zaro aloqador axborotni baholash (Mutual informatsion) metodlari orqali aniqlash mumkinligi taklif sifatiga ko'rsatilgan. Maqolada **oronims.uz** – web ilovasi o'zbek milliy korpusidagi oronimlar bazasini yaratish muammosi o'rta tashlangan.

Abstract. In the article, the science of the 21st century allows extensive research on the example of electronic corpora, focusing on the issue of the national linguistic nature of Oronim, their lexicography, and their inclusion in the national corpus. The problem of their reflection in a single national corpus is highlighted due to the scientific research carried out in the field of toponymy, the concept of oronims and oronymics in the sources, as well as meeting different opinions on the issue of defining the object of its study. Currently, the possibility of automatic identification of phonemes through latent-semantic analysis (Latent semantic analysis), parallel text alignment (Bitext word alignment) and mutual information evaluation (Mutual informative) methods has been shown to the quality of the proposal. The article highlights the problem of creating a database of oronims in the Uzbek national corpus of oronims.uz - a web application.

Аннотация. В статье наука XXI века позволяет проводить масштабные исследования на примере электронных корпусов, уделяя особое внимание вопросу национальной языковой природы оронимов, их лексикографии, включения в национальный корпус. Проблема их отражения в едином национальном корпусе освещается благодаря научным исследованиям, проведенным в области топонимики, понятию оронимов и оронимики в источниках, а также встрече различных мнений по вопросу определения объекта ее изучать. В настоящее время для качества предложения показана возможность автоматической идентификации фонем методами латентно-семантического анализа (Latent Semantic Analysis), параллельного

выравнивания текста (Bitekt word выравнивание) и оценки взаимной информации (Mutual информативный). В статье освещена проблема создания базы данных оронимов в узбекском национальном корпусе oronims.uz – веб-приложении.

Kalit soʻzlar: *oronim, toponim, gidronim, speleonim, relef, korpus lingvistikasi, latent-semantik tahlil, kollokatsiya, parallel matnlarni tenglashtirish, latent-semantik tahlil, web ilova.*

Kirish

Jahon tilshunosligida onomastika, xususan, toponimiyaning bir turi sifatida oronimiya ham xalq tarixi va til tarixi bilan chambarchas bogʻliq boʻlib, milliy til, xalqning iqtisodiy-ijtimoiy hayoti, madaniy-maʼnaviy munosabatlarining oʻziga xos xususiyatlarini aks ettiruvchi lisoniy manba hisoblanadi. Tilning sistematiy birligi boʻlgan mazkur lingvomadaniy birliklar tadqiqi ayniqsa, leksikologiya, leksikografiya, korpus lingvistikasida tez-tez koʻzga tashlanmoqda: XXI asr ilmi-fani bu borada elektron korpuslar misolida keng tadqiq imkonini yaratgan boʻlsa-da, oronimlarning milliy lisoniy tabiati ularn lugʻatlashtirish hamda milliy korpusda berish ishlarini bir qadar murakkablashtiradi. Binobarin, Oʻzbekiston toponimik tizimidagi oronimik qurilmalarni qayta ishlanuvchi, lingvistik tasniflangan, mashinada oʻqiladigan metalingvistik maʼlumotlar formatida taqdim etish muammosi hanz dolzarbligicha qolmoqda.

Oʻzbek tilshunosligida toponimikaga oid ilmiy tadqiqotlarda oronimika tushunchasi, uning qamrovi va bu soha oʻrganadigan masalalar koʻlami aniq belgilanmagan va koʻrsatilmagan. Ilmiy adabiyotlarda bu masalaga doir quyidagi izohlarni kuzatish mumkin: “Oronimika – toponimikaning maʼlum bir til yoki region oronimlari majmuyini oʻrganadigan boʻlimi” [Djʼsimov, Tillaeva, 2002]; “Oronimika – toponimikaning oronimlarni vujudga kelishi, taraqqiyot qonunlarini va amal qilishini oʻrganadiga kichik boʻlimi” [Korayev, 2005]. E.Begmatov va N.Uluqovlar esa oronimikaga “Toponimikaning oronimlarni paydo boʻlishi va vazifaviy xususiyatlarini oʻrganuvchi tarmogʻi” deb izoh beradilar [Begmatov, Uluqov].

Yuqorida keltirilgan fikrlarni umumlashtirgan holda shuni taʼkidlash oʻrinliki, toponimiyaga oid ilmiy adabiyotlar va ilmiy tadqiqotlarda oronim va oronimiya tushunchalariga nisbatan quyidagi talqinlar uchraydi: 1) faqat togʻ (relefning musbat shakllari) nomlari; 2) relefning musbat va manfiy shakllari nomlari; 3) relefning musbat va manfiy hamda oraliq (nol) shakllari nomlari.

Oronim termini asli leksik jihatdan “togʻ nomi” maʼnosini anglatga-da, lekin uning maʼno doirasi ancha keng. Oronimlarga togʻlar, togʻ tizmalari, qirlar, dovonlar, tepalar, balandliklar, qoyalar, choʻqqilar, toshlar, adirliklar va togʻliklarda joylashgan boshqa musbat xarakterdagi orografik obyektlarning atoqli otlari bilan bir qatorda, oʻzining qaysi bir jihati bilan shu obyektlar bilan chambarchas bogʻliq

bo'lgan dara va uning turli ko'rinishlari, jar va uning turli ko'rinishlarini o'z ichiga oluvchi relefning manfiy shakllari nomlarini kiritish maqsadga mubofiq bo'ladi.

Oronimika onomastikaning ilmiy yo'nalishi sifatida tilshunoslikning dialektologiya, etimologiya, terminologiya, morfemika, til tarixi kabi bo'limlari bilan uzviy aloqador.

Hozirda oronimlarni avtomatik aniqlashning quyidagi metodlari orqali aniqlash mumkin:

1. Latent-semantik tahlil (Latent semantic analysis).
2. Parallel matnlarni tenglashtirish (Bitext word alignment).
3. O'zaro aloqador axborotni baholash (Mutual informatsion)[15].

Qo'llanilgan dasturlash tillari:

- PHP v7.2 (Laravel Framework v6.2)
- HTML
- SASS/CSS (Bootstrap v4)
- Javascript (Vuejs v2.5, Jqueru v3.4)

Qo'llanilgan ma'lumotlar bazasi:

- MuSQL v5.5

(Dasturning kod qismidan ba'zi lavhalar:

oronimController.php:

```
<?php
```

```
namespace App\Http\Controllers;
```

```
use App\Example;use App\oronim;
```

```
use App\oronimLang;
```

```
use Illuminate\Http\Request;
```

```
class oronimController extends Controller{public function__construct()  
{ $this->middleware('auth')->except(['index','search','show']); } }
```

An'anaga ko'ra parallel matnlarni birlashtirish uchun statistik usullardan foydalaniladi. Ular bir tildagi matnlar uzunligi va boshqa tildagi matnlar uzunligi o'rtasida so'zlar yoki belgilar bilan o'lchanadigan taxminiy muvofiqlikning mavjudligiga asoslanadi. Shuning uchun, bitta tildagi matnlarning uzunligiga qarab, boshqa tilda matnning mos parchalari qayerdan boshlanishi kerakligini hisoblash mumkin. Faqatgina nisbatan ko'proq mos keladigan muvofiqlikni tanlash qoladi. Ushbu usullarfaqat so'zlar soni yoki belgilar sonini hisoblashga asoslangani uchun tezkor usullar hisoblanadi.

Qayd etish kerakki, Facebook, Pinterest, Twitter ijtimoiy tarmoqlari semantik razmetkasiz ish ko'rmaydi. Jumladan, ushbu web-xizmatlar mavjud maqolalarning bir qismini oladi va ularni o'z platformalarida namoyish etadi. Agar biz semantik HTML dan foydalansak, ular bu ishni yaxshiroq, sifatliroq bajarishlari mumkin.

Semantik HTML beb-saytlarga kirish imkonini oshiradi.

Ekrandan o'qish imkonini beruvchi kabi yordamchi texnologiyalar mavjud HTMLni tahlil qiladi va sharhlaydi (talqin qiladi). Semantik HTML yordamida maxsus ehtiyojlari bo'lgan odamlar tegishli maqolalarni o'qishlari va yo'nalish olishlari osonroq bo'ladi. Biroq bu hammasi emas. HTMLni tushunishga harakat qiladigan son-sanoqsiz boshqa mashinalar ham mavjud. Ular internetning katta qismini tashkil qiladi va ularni yanada mazmunli ma'lumotlar bilan ta'minlash katta hajmli mehnatni talab qiladi.

Bu taxminan quyidagicha manzaraga ega:

```
<!DOCTYPE html>
<html itemscope itemtype="https://schema.org/Article" lang="" dir="">
<head>
<title itemprop="name"></title>
<meta itemprop="description" name="description" content="">
<meta itemprop="author" name="author" content="">
</head>
<body>
<article>
<header>
<h1 itemprop="headline"></h1>
<time itemprop="datePublished" datetime=""></time>
<p><a itemprop="author" href=""></a></p>
</header>
<div itemprop="about"></div>
<div itemprop="articleBodu">
<!-- maqolaning asosiy matni bu yerda bo'lishi kerak-->
</div>
<footer>
<time itemprop="dateModified" datetime=""></time>
<section itemscope itemtype="http://schema.org/WebPage">
<!-- Sarlabha-->
<h2></h2>
<p><a itemprop="relatedLink" href=""></a></p>
</section>
</footer>
</article>
</body>
</html>
```

Orografik obyektlarni ifodalovchi atamalar va oronimlarning tarkibida indikator vazifasida qo'llanuvchi quyidagi so'zlarni ko'rsatish mumkin: tog', tosh, qoya, qiya, cho'qqi, qir, tepa, dobon, bel, sirt, adir, ko'tal, ort, do'ng, cho'l, gaza, dara, tangi, zob, jar, ko'h, kamar, to'da, sang, dasht, mo'la, nura, o'r, o'ra, maydon, aydar, mas//mos, taxta, alang, qo'l, shoti, ayri, suyri, aqba//ag'ba, yaylov, yayloq,

chobra, bitob//pitob, xuchchi, tubak, qaloq, do'm, bo'ktar, zayak, hayat, qashqa, qat, qatob, qoboq, singir, sala, jun//jo'na, to'qay, chag'il//chag'al, chag'at, supa, yona, shiram, chuqur, xandaq, g'urum, qum, chamber, qaqir, taqir, dagar, jol, to'la, bayir, sho'r, bet, habol, g'or, poya, qo'ruq, tuz, chalqa, tumshuq, qapug', darboza, darband, qiz//quz, ti//tig'//teg', baroz//barzu, shirdon, bo'z, jira, qayir, bo'yin, bo'ynoq, boldir, changal//jangan, pushta, qasoba//qasag'a, qasnaq, orqa, orqabul, burun, oshub, qayqi, o'y, toqcha, qapchig'ay, yobon, xobar, jobuz, zarang, sahra, chortoq, sangir, bodiya, mag'ora, barg'//barq, dapsan, tarma, tirsak, chosh//shosh, chek, ungur, chordara, chig'anaq va b.

Shablonning HTML-razmetkasi article, header, footer singari semantik elementlardan iborat.

Bilamizki, korpus lingvistikasi:

1. Nutq tadqiqiga doir natija va xulosalarni aniqlash;
2. Empirik nutqiy materialning qamroviga oid tizimli va keng ko'lamli lingvistik tadqiqotlarni amalga oshirishni maqsad qiladi.

Shunga ko'ra korpusda qamralayotgan material birliklarining razmetkasi lingvistik va tashqi lingvistik turlariga bo'linadi. Bugungi kunda razmetkaning standartlashgan vositalarini ishlab chiqishga mo'ljallangan maxsus xalqaro Text Encoding Initiative (TEI) loyihasi amalga oshirilgan bo'lib, bunda elektron hujjatlarning razmetkasi uchun SGML xalqaro razmetka tili yoxud XMLning turlari qo'llaniladi.

Oronimik konversiya usulida hosil bo'lgan oronimlarni ikki guruhga ajratish mumkin:

1. Tub asosdan hosil bo'lgan oronimlar: a) orfografik terminlar oronimlashadi: Ag'ba (Qm., dovon), Kamar (tog'li hududlarda, joy), Gaza (Bs., joy), Kachar (G'z., adir), Sangar (Mk., yaylov), Tarpi (Chr., balandlik), Taqir (Do., balandlik); Ko'tal (Chr., tog', q-q.), Chimgan (G'z., adir), Qayg'i (Bs., adir), Qasaba (Do., balandlik), G'urum (Do., balandlik), G'urim (Do., balandlik), Hasar (Chr., balandlik); b) etnonimlar oronimlashadi: Barmoq (Qm., yaylov, balandlik), Lochin (Shs., cho'qqi), Manqa (Do., adir), O'lmas (Ko'kabuloq, tog'), Uyqur (Do., dara); b) gidrografik terminlar oronimlashadi: Kariz (Kt., adir), Orna (G'z., jar).

2. Yasama so'zlardan hosil bo'lgan oronimlar. Bu guruhga kiruvchi oronimlarni tarkibiy tuzilish xususiyatiga ko'ra yana ikki ichki guruhga ajratish mumkin: a) sodda tarkibli yasama so'zlardan hosil bo'lgan oronimlar: Abrikli (CHr., adir), Andizli (Chr., Qm., Bs., dara), Arnali (G'z., adir), Ahalli (Kt., adir), Bedak (Kt., joy), Burg'onli (Yab., adir), Dubolli (Bs., dara), Jo'jali (Do., dara), Irg'ayli (Qm., adir), Isbanni (Qm., Yab., yaylov, joy), Kapali (G'z., SHs., Mb., yaylov, dasht), Ko'hak (Qo'ng'irtog'ning qadimgi nomi), Surxak (Kt., balandlik), Xabdonli (G'z., yaylov), Chaqishli (SHo., adir), Cho'yanli (YAv., Bs., adir), Sherak (SHo., adir), O'rdali (Jq., tepa), Qazma (G'z., yaylov), Qayroqli (Sho., balandlik), Qarg'ali (Shs., tepa; So., adir), Qarg'anli (Shs., adir) ba b.;

b) qo'shma tarkibli so'zlardan hosil bo'lgan oronimlar: Abg'ontaxta (Do., balandlik), Ajinazob (Qm., balandlik), Ayg'irtepa (G'z., Do., Bs., tepa), Aylanatumshug' (Kt., balandlik), Archamaydon (Shs., tog'), Asqartepa (Yab., tepa), Asqarcho'qqi (Yab., cho'qqi), Baland qo'ton (Chr., adir), Bag'ishtepa (Os., tepa), Beshqiz (Qm., tepa), Beshqo'l (Do., dara), Boyirgaza (Chr., gaza), Bo'rijar (Shs., jar), Bo'g'masoy (Do., dara), Go'rdara (Do., dara), Do'g'irtosh (Dn., balandlik), Etimtaxta (Do., tog') va b.;

b) izohli birikmalardan hosil bo'lgan oronimlar: Bog'ibaland, Go'rimor, Go'riqalmoq, Darayiturko, Darixitoy, Dashtikalta, Dashtichinor, Dahnidarg'oz, Zaxakimaron, Kamariturk, Kamarichiljubut, Ko'hisafed, Ko'hisiyoh, Ko'hichosh, Xamikalon, Chashmiganda, Shibarikalon va b.

g) predikativ birikmalardan hosil bo'lgan oronimlar: Xontushdi (tepa), Lo'limurda (adir), Shuturmurda (dara), Xarduzd (tog'), Darg'amurda (adir), Asppar (dara), Xamixukkushtagi (adir), Ayiquchdi (dara).

Biz taklif qilayotgan milliy oronimlar bazasining dasturiy ta'minoti quyidagilarni o'z ichiga oladi: **oronims.uz** – web ilovasi o'zbek milliy korpusidagi oronimlar bazasini tashkil etish va undan qulay foydalanish imkonini beruvchi tizim hisoblanadi. U o'zida oronimlar bazasini to'ldirib borish va so'ralgan oronimni tushunarli va qulay tarzda batafsil ma'lumotlar asosida taqdim etish funksiyalarini mujassam etgan. Ilova tarmoqqa (internet) ulangan server vazifasini o'tay oladigan kompyuterda ishlashga mo'ljallangan. Unga HTTP protokoli orqali murojaatlar yuborilishi va javob olinishi ko'zda tutilgan. Barcha web ilova (sayt)lar kabi **oronims.uz** web-ilovasi ham o'zi o'rnatilgan server kompyuterda web-server paketlari ishlashini taqozo etadi. Ushbu bazaning texnik bosqichini yaratishda quyidagi dasturlash tillari tanlandi.

Dastur turi: Web ilova (web application).

Qo'llanilgan dasturlash tillari:

PHP v7.2 (Laravel Framework 6.2) HTML SASS/CSS (Bootstrap v4)
Javascript (Vuejs v2.5, Jqueru v3.4)

Qo'llanilgan ma'lumotlar bazasi: MySQL v5.5

Xulosa

Xulosa o'rnida shuni aytish joizki, til korpusidagi oronimlar bazasi ta'limga, ayniqsa, til ta'limiga doir quyidagi tipik xatolarni bartaraf etish imkonini beradi. Jumladan, taqdim etilgan so'zlar qatoridan so'zlarni to'g'ri tanlashni bilmaslik; oronimik birliklarni bilmaslik; mashhur oronimlardan bexabarlik.

Foydalanilgan adabiyotlar:

1. Авлакулов Я.И. Ўзбек тили ономастик бирликларининг лисоний тадқиқи: Филол. фан. номз... диссер автореф. – Тошкент, 2012. – Б. 27.
2. Бегматов Э., Улуков Н. Ўзбек ономастикаси терминларининг изоҳли луғати. – Наманган, 2006. – Б. 63.
3. Дўсимов З., Тиллаева М. Топонимика асослари. – Тошкент, 2002. – Б. 82.
4. Исломов И. Ўзбек тилида ер сатҳи кўринишларини ифодаловчи атов бирликлар тизими: Филол. фан. фалсафа докори (PhD) диссертацияси. – Қарши, 2019. – 124 б.
5. Суперанская А.В. Что такое топонимика? – Москва: Наука, 1984. – С. 22.
6. Подольская Н.В. Словарь русской ономастической терминологии. – Москва: Наука, 1978. – С. 105.
7. Пospelов Е. М. Оронимика СССР: состояние и проблемы// Оронимика (Сборник статей). – Москва, 1969. С. 4
8. Темиров Ш. Самарқанд вилояти оронимлакрининг лисоний тадқиқи: Филол. фан. ном. ... дисс. автореф. – Самарқанд, 2019. – 24 б.
9. Каримов С.А., Бўриев С.Н. Ўзбек топонимикаси тараққиёт босқичида. – Тошкент: Фан, 2006. -Б. 44.
10. Корпусная лингвистика как раздел языкознания. — Текст : электронный // MyfilologY.ru – информационный филологический ресурс : [сайт]. – URL: <https://myfilology.ru/177/korpusnaya-lingvistika-kak-razdel-yazykoznanieYa/> (дата обращения: 5.02.2021)
11. Қораев С. Географик номлар маъноси. – Тошкент: Ўзбекистон, 1978. – Б. 193.
12. Qorayev S. Toponimika. – Toshkent: O'zbekiston faylasuflari milliy jamiyati, 2006. – B. 66.
13. Қораев С. Ўзбекистон вилоятлари топонимлари. – Тошкент: Ўзбекистон миллий энциклопедияси, 2005. – Б. 217.
14. Эназаров Т. Номшунослик масалалари. – Тошкент, 2010. – Б.78.
15. Шильнов, А. Г., Миллер, Д., Митчелл, П. Д., & Смокотин, В. М. (2020). Сравнительный анализ идиом, представленных в российских Учебниках английского языка, и идиом, представленных в современных аутентичных источниках. Язык и культура, (49), 157-174. <https://doi.org/10.17223/19996195/49/10>
16. Шоабдирахмонов Ш., Асқарова М. ва б. Ҳозирги ўзбек адабий тили. – Тошкент, 1980. – Б. 182.
17. Ўзбек тили грамматикаси. 1-қисм. – Тошкент, 1975. – Б.34.
18. Toirova G. The Role Of Setting In Linguistic Modeling. International Multilingual Journal of Science and Technology. ISSN: 2528-9810 Vol. 4 Issue 9, September – 2019, -P.722-723



19. Toirova, G. Creation and importance of language corps in Uzbekistan. BIO Web of Conferences, 2024, 84, 04003.
20. Toirova G., Astanova G., Rahimova N. Artistic Expressions of a Situational Pragmatic SYstem. // International Journal of Recent TechnologY and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019. –P.4591-4593
21. Toirova G., Yuldasheva M., Elibaeva I. Importance of Interface in Creating Corpus. // International Journal of Recent TechnologY and Engineering, 2019, 8(2 Special Issue 10), P. 352–355
22. Toirova G, Abdurahmonova N., Ismoilov A., Applying Web Crawler Technologies for Compiling Parallel Corpora as one Stage of Natural Language Processing. // Proceedings - 7th International Conference on Computer Science and Engineering, UBMK 2022, 2022, P. 73–75
23. O'zbek Milliy Korpusida idiomalar bazasi deponentlanadigan materiallari. Huquq egasi Begmatova Gulchehra Xolmurodovna. Mualliflar: Begmatova Gulchehra Xolmurodovna. Ulug'ov Anvar Abduraimovich. – 2019.