

MUALLIFLIK KORPUSI QIDIRUV TIZIMINI ISHLAB CHIQISH

Sanjarbek Baxodirov Rahmatali o'g'li,
Kompyuter lingvistikasi mutaxassisligi magistranti
sanjarbahodirov9901@gmail.com
ToshDO'TAU

Nilufar Muradova Baxodir qizi,
Kompyuter lingvistikasi mutaxassisligi magistranti
muradovanilufar06@gmail.com
ToshDO'TAU

Annotatsiya. Mualliflik korpusi qidiruv tizimi zamonaviy lingvistik tadqiqotlar uchun samarali vosita hisoblanadi. Ushbu tizim foydalanuvchilarga matnlarni konkordans qidiruv va kengaytirilgan qidiruv yordamida tahlil qilish imkonini beradi. Konkordans qidiruv so'z shakllari va ularning kontekstlarini aniqlashda yordam bersa, kengaytirilgan qidiruv turli filtrlash parametrlari orqali aniq natijalar taqdim etadi. Mazkur maqolada ushbu qidiruv tizimining asosiy tamoyillari, ishlash mexanizmi va texnik jihatlar yoritiladi. Tizimning samaradorligi va foydalanuvchilarga qulayligi haqidagi tahlillar ham keltiriladi. Mualliflik korpusining ushbu qidiruv imkoniyatlari lingvistika, tarjima va matnshunoslik sohalari uchun muhim ahamiyat kasb etadi.

Abstract. The authorship corpus search system is an effective tool for modern linguistic research. This system allows users to analyze texts through concordance search and advanced search functionalities. The concordance search helps identify word forms and their contexts, while the advanced search provides precise results through various filtering parameters. This article highlights the fundamental principles, operational mechanisms, and technical aspects of this search system. It also includes an analysis of the system's efficiency and user-friendliness. The search capabilities of the authorship corpus are of great importance for linguistics, translation, and text studies.

Аннотация. Система поиска по авторскому корпусу является эффективным инструментом для современных лингвистических исследований. Данная система позволяет пользователям анализировать тексты с помощью поиска по конкордансам и расширенного поиска. Поиск по конкордансам помогает определить формы слов и их контексты, а расширенный поиск предоставляет точные результаты благодаря различным параметрам фильтрации. В данной статье рассматриваются основные принципы, механизм работы и технические аспекты данной поисковой системы. Также приводится анализ её эффективности и удобства использования. Возможности поиска в авторском корпусе имеют важное значение для лингвистики, перевода и текстоведения.

Kalit soʻzlar: *Mualliflik korpusi, konkordans qidiruv, kengaytirilgan qidiruv, lingvistik tahlil, matnshunoslik, qidiruv algoritmlari, korpus lingvistikasi.*

KIRISH

Hozirgi kunda lingvistik tadqiqotlar sohasida korpus lingvistikasi muhim ahamiyat kasb etmoqda. Matnlarni chuqur tahlil qilish, leksik va grammatik xususiyatlarini oʻrganish, tarjima jarayonlarini soddalashtirish va sunʼiy intellekt tizimlarini rivojlantirish uchun korpuslardan foydalanish samarali usullardan biri hisoblanadi. Shu nuqtai nazardan, mualliflik korpusi maʼlumotlarni tizimli ravishda toʻplash, saqlash va tahlil qilish imkonini beruvchi qudratli vositadir.

Mualliflik korpusi yordamida matnlar tarkibiy qismlarini oʻrganish, soʻzlarning ishlatilish chastotasini aniqlash, semantik bogʻliqliklarni tadqiq qilish hamda tarjima amaliyotida muhim boʻlgan parallel matnlarni tahlil qilish imkoniyati yaratiladi. Ushbu maqolada mualliflik korpusi uchun ishlab chiqilgan qidiruv tizimining ishlash tamoyillari, asosiy funksiyalari va texnik jihatlari koʻrib chiqiladi.

Mazkur tadqiqotda korpus asosida ishlovchi qidiruv tizimlarining samaradorligi va foydalanuvchilarga qulayligi tahlil qilinadi. Shu bilan birga, lingvistik tadqiqotlarda ushbu tizimdan foydalanish imkoniyatlari, kelajakda uni yanada takomillashtirish yoʻnalishlari muhokama qilinadi. Ushbu maqola nafaqat lingvistlar, balki tarjimonlar, tadqiqotchilar va sunʼiy intellekt tizimlari bilan shugʻullanuvchilar uchun ham foydali boʻlishi mumkin.

ASOSIY QISM

Konkordans qidiruvi. Matnlarni lingvistik jihatdan tahlil qilishda soʻzlarning kontekstini oʻrganish muhim vazifalardan biri hisoblanadi. Shu sababli, bizning qidiruv tizimimizning birinchi qismi Konkordans qidiruvi boʻlib, u maʼlum bir soʻz yoki iborani uning kontekstida izlash imkonini beradi. Ushbu tizimni Python dasturlash tili va Django veb-frameworkidan foydalangan holda koʻrib chiqildi.

Konkordans qidiruv tizimi uchta asosiy qism dan iborat:

1. **Maʼlumotlar bazasi modeli** – Konkordans jadvali soʻzlarni va ularning kontekstini saqlaydi.
2. **Qidiruv funksiyasi** – `concordance_search` funksiyasi foydalanuvchining kiritgan soʻzi boʻyicha mos keluvchi natijalarni chiqaradi.
3. **Maʼlumotlarni indekslash** – `build_concordance` kommandasi yordamida matnlardagi soʻzlar avtomatik ravishda **Concordance** bazasiga yuklanadi.

Konkordans modeli (maʼlumotlar bazasi modeli). Konkordans qidiruvi asosiy maʼlumotlar bazasi modeli sifatida **Concordance** modelidan foydalanadi.

Ushbu model har bir soʻzning kontekstini, uning joylashuvi va tegishli asar bilan bogʻliqligini saqlaydi.

```
class Concordance(models.Model):
    work = models.ForeignKey(Work, on_delete=models.CASCADE)
    keyword = models.CharField(max_length=100)
    context_before = models.TextField()
    context_after = models.TextField()
    position = models.IntegerField()

    def __str__(self):
        return f"{self.keyword} in {self.work.title}"
```

Ushbu model quyidagi maydonlar(fields)dan iborat:(Qarang:1-jadval)

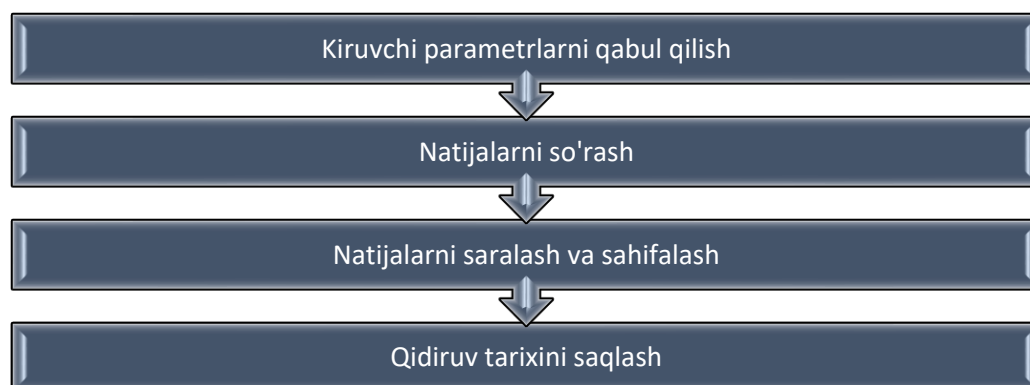
1-jadval. Konkordans qidiruv modeli tuzilishi

Maydon	Turi	Izoh
id	AutoField (PK)	Unikal identifikator
work_id	ForeignKey(Work)	Ushbu soʻz qaysi matnga tegishli ekanligi
keyword	CharField(100)	Qidirilayotgan soʻz
context_before	TextField()	Soʻzdan oldingi kontekst
context_after	TextField()	Soʻzdan keyingi kontekst
position	IntegerField()	Soʻzning matndagi tartib raqami

2. *Concordance qidiruv funksiyasi* – matnda soʻzlarni kontekst (oldingi va keyingi soʻzlar) bilan birga izlash tizimi. Ushbu usul lingvistik tahlil, matnni semantik tekshirish, maʼlumot qidirish va tilshunoslik sohalarida keng qoʻllaniladi.

Django yordamida qurilgan ushbu qidiruv tizimi maʼlum matnlardan soʻzlarni chiqarib, ularning atrofidagi kontekst bilan birga saqlaydi va qidiruv imkoniyatlarini taqdim etadi.

concordance_search funksiyasi quyidagi bosqichlarda ishlaydi:



Kiruvchi Parametrlarni Qabul Qilish. Foydalanuvchi qidiruv soʻzini va qoʻshimcha filtrlarni (janr, davr, qidiruv turi) kiritadi:

```
query = request.GET.get('q', "")
genre = request.GET.get('genre')
period = request.GET.get('period')
search_type = request.GET.get('search_type', 'exact')
```

Qidiruv turlari:

1. exact – aynan mos keluvchi soʻzlarni qidirish (keyword__iexact=query)
2. starts_with – soʻz boshi boʻyicha qidirish (keyword__istartswith=query)
3. contains – soʻz tarkibida berilgan qator bor-yoʻqligini tekshirish (keyword__icontains=query)

Natijalarni Soʻrash. Qidiruv soʻroviga mos keluvchi natijalar Concordance modelidan olinadi:

```
if search_type == 'starts_with':
    results = Concordance.objects.filter(keyword__istartswith=query)
elif search_type == 'exact':
    results = Concordance.objects.filter(keyword__iexact=query)
else:
    results = Concordance.objects.filter(keyword__icontains=query)
```

Agar foydalanuvchi janr yoki davr boʻyicha filtr qoʻllasa, natijalar quyidagicha filtrlanadi:

```
if genre:
    results = results.filter(work__genre_id=genre)
if period:
    results = results.filter(work__period_id=period)
```

Natijalarni Saralash va Sahifalash. Topilgan natijalar *asar nomi* va *soʻzning tartib raqami* boʻyicha saralanadi:

```
results = results.order_by('work__title', 'position')
```

Natijalar sahifalarga boʻlinadi (20 tadan):

```
results_paginator = Paginator(results, 20)
concordances = results_paginator.page(results_page)
```

Qidiruv Tarixini Saqlash. Qidiruv tizimi foydalanuvchi kiritgan soʻrovlarni tarixga yozib boradi:

```
SearchHistory.objects.create(  
    query=query,  
    results_count=total_results,  
    filters_used={'genre': genre, 'period': period, 'search_type': search_type}  
)
```

Ma'lumotlarni Indeksflash. Ma'lumotlarni indekslash – konkordans qidiruv tizimining muhim qismi bo'lib, u matnlardagi so'zlarni ajratib, ularning kontekstini aniqlash va ma'lumotlar bazasiga yuklash jarayonidir. Bu jarayon konkordans qidiruvning samarali ishlashi uchun zarur bo'lib, foydalanuvchilar so'zlarni qidirganda ularni o'z konteksti bilan birga topish imkonini beradi.

Indeksflash jarayoni quyidagi bosqichlardan iborat:

Birinchii bosqichda barcha matnlar Work modelidan olinadi. Ushbu model ma'lumotlar bazasida saqlanayotgan asar yoki hujjatlarni ifodalaydi. Indeksflash funksiyasi har bir matn ustida ishlaydi va uni tahlil qiladi.

Ikkinchi bosqichda matn so'zlarga bo'linadi. Buning uchun maxsus matnni tahlil qilish usullari ishlatiladi. Masalan, regular ifodalar (regex) yordamida barcha so'zlar ajratib olinadi va ular alohida ro'yxatga joylashtiriladi. Har bir so'z o'zining matndagi joylashuv tartibiga ega bo'ladi.

```
for i, work in enumerate(works, 1):  
    self.stdout.write(f'Processing work {i}/{total_works}: {work.title}')  
  
# 3. Matnni kichik harflarga o'tkazib, so'zlarni ajratish  
words = re.findall(r'\b\w+\b', work.content.lower())
```

Uchinchi bosqichda har bir so'zning konteksti aniqlanadi. Bu bosqichda har bir so'zning oldidan va keyin keluvchi besh-oltita so'z ajratib olinadi. Agar so'z matnning boshida yoki oxirida joylashgan bo'lsa, u holda mavjud bo'lgan qismi olinadi. Kontekstni aniqlash natijasida qidiruv tizimi foydalanuvchi izlayotgan so'z atrofidagi ma'lumotni ham taqdim etadi.

```
for position, word in enumerate(words):  
    if len(word) < 2: # Bitta harfdan iborat so'zlarni o'tkazib yuborish  
        continue
```

```
# 4. So'zning oldidan va keyin keluvchi so'zlarni aniqlash  
start_pos = max(0, position - 5)  
end_pos = min(len(words), position + 6)
```

```
context_before = ''.join(words[start_pos:position])  
context_after = ''.join(words[position+1:end_pos])
```

To'rtinchi bosqichda so'z va uning konteksti Concordance jadvaliga yoziladi. Ushbu jarayonda har bir so'zga tegishli quyidagi ma'lumotlar bazaga qo'shiladi:

- So'zning o'zi

- Matndagi joylashuv tartibi
- Oldingi soʻzlar (kontekst oldin)
- Keyingi soʻzlar (kontekst keyin)
- Ushbu soʻz tegishli boʻlgan asar identifikatori

Barcha matnlar tahlil qilinib, kerakli maʼlumotlar bazaga yuklangandan soʻng, konkordans qidiruv tizimi ishlashga tayyor boʻladi. Endilikda foydalanuvchi soʻzlarni izlaganda, ularning matndagi aniq joylashuvi va konteksti bilan birga koʻrish imkoniyatiga ega boʻladi.

Indekslash jarayoni muntazam ravishda yoki yangi matnlar qoʻshilganda qayta ishga tushirilishi mumkin. Buning uchun maxsus Django management command yoziladi, bu buyruq chaqirilganda barcha mavjud matnlar qayta tahlil qilinadi va konkordans jadvali yangilanadi.

Umuman olganda, indekslash jarayoni konkordans qidiruv tizimining asosiy tayanch qismi boʻlib, u soʻzlarni samarali qidirish va kontekstda koʻrsatish imkonini beradi.

NATIJA. Konkordans qidiruv tizimi matn tahlili va lingvistik tadqiqotlar uchun muhim vositalardan biri boʻlib, matndagi soʻzlarni ularning konteksti bilan birga topish imkonini beradi. Ushbu tadqiqot natijasida concordance qidiruvi amalga oshirishning samarali usullari ishlab chiqildi va tizimning asosiy ishlash tamoyillari aniqlandi.

Tadqiqot davomida matnlarni toʻliq tahlil qilish va qidiruv jarayonini tezlashtirish uchun maxsus indekslash mexanizmi yaratildi. Ushbu indekslash jarayonida matndagi har bir soʻz ajratib olinib, ularning joylashuvi va konteksti qayd etildi. Bu esa qidiruv natijalarini yanada aniqroq va samaraliroq qilish imkonini berdi. Foydalanuvchilar kiritgan soʻzni nafaqat mustaqil holda, balki uning oldidan va keyin keluvchi soʻzlar bilan birgalikda topish imkoniyatiga ega boʻldilar.(Qarang:1-rasm)

Qidiruv natijalari matndagi soʻzlarning qanday ishlatilishini tushunish, uslubiy tahlillar oʻtkazish va lingvistik tadqiqotlar olib borish uchun muhim ahamiyat kasb etdi. Tizim real vaqt rejimida tezkor natijalar taqdim etib, katta hajmdagi matnlarni samarali qayta ishlashga yordam berdi. Shuningdek, soʻz shakllari, sinonimlar va matn strukturasi bilan ishlash imkoniyatlari ham qoʻshildi, bu esa qidiruvning aniqroq boʻlishiga zamin yaratdi.

Abdurauf Fitrat korpusi Asosiy Haqida **Konkordans** Kengaytirilgan qidiruv Asarlar Lemmalar Til statistikasi

Konkordans qidiruv

So'z shakli topish

☒ Aniq yozilishi ☐ So'z boshlanishi ☐ So'zning istalgan qismi

Janr: Barcha janrlar

Qidirish

Umumiy asarlar bo'yicha

A (3023)	B (4835)	D (1597)	E (1329)	F (921)	G (967)	H (1759)	I (2059)	J (925)	K (3932)	L (539)	M (3628)	N (1433)	O (4664)	P (748)	Q (4381)	R (1007)	S (4291)
T (5379)	U (1581)	V (600)	X (1434)	Y (3651)	Z (733)	O' (2070)	G' (305)	Sh (1124)	Ch (1262)								

Natijalar (73 ta)

kim gulshan ichra gul ochilg'ay **gul** uzra turmay sunbul ochilg'ay v

[Aruz haqida](#) [Risola](#)

1-rasm. Konkordans qidiruv natijasi

XULOSA

Ushbu tadqiqot natijasida concordance qidiruv tizimi matn tahlili uchun muhim vosita ekanligi tasdiqlandi. Tizim matndagi har bir so'zning joylashuvi va kontekstini aniqlashga imkon berib, qidiruv natijalarini aniq va tushunarli shaklda taqdim etadi.

Indekslash jarayoni orqali matn tuzilmasi chuqur tahlil qilinib, so'zlarning qaysi kontekstda ishlatilganini aniqlash imkoni yaratildi. Ushbu yondashuv lingvistik tadqiqotlar, matnni chuqur o'rganish va turli tahliliy ishlarda samarali natijalar berishi mumkin.

Natijalar shuni ko'rsatadiki, konkordans qidiruvi lingvistik izlanishlar, tarjima ishlari, matn muallifligini aniqlash, adabiy tahlillar va axborot qidiruv tizimlarida keng qo'llanilishi mumkin bo'lgan kuchli vosita ekanligi isbotlandi. Kelgusida tizimni yanada rivojlantirish uchun tabiiy tilni qayta ishlash (NLP) usullaridan foydalangan holda lemmatizatsiya, ma'noli tahlil va kontekstga asoslangan chuqur qidiruv imkoniyatlarini qo'shish rejalashtirilmoqda.

Foydalanilgan adabiyotlar:

1. Hamroyeva Sh. O'zbek tili mualliflik korpusini tuzishning lingvistik asoslari: Monografiya. – Globe edit, 2020.



2. Hamroyeva Sh. Korpus lingvistikasi atamalarining qisqacha izohli lug'ati: Terminologik lug'at. – Globe edit, 2020.
3. Захаров В.П. Корпусная лингвистика. Учебно-методическое пособие. – Санкт-Петербург, 2005.
4. <https://uzlc.tsuull.uz/index.php/uzlangcult/article/download/149/151/473>
5. <http://compling.navoiy-uni.uz/index.php/conferences/article/view/1038>