SKETCHE ENGINE KORPUS MENEJERI VA UNING AYRIM IMKONIYATLARI

Rizvanov Qodir Qaxramon o‘g‘li
rizvanovqodir@gmail.com
Alisher Navoiy nomidagi
Toshkent davlat o‘zbek tili va adabiyoti universiteti
Kompyuter lingvistikasi
mutaxassisligi 1-kurs magistranti
+998971223377

Annotatsiya. Ushbu maqolada Sketche Engine korpus menejerining imkoniyatlari keltirigan. Ushbu korpus menejerining dasturiy ta’minot sifatidagi xususiyatlari, undan foydalanishning ayrim qoidalari, Sketche Enginening talabalar, ilmiy izlanuvchilar, tarjimon va tilshunoslar uchun ahamiyati tavsiflangan.

Abstact. In this article description the capabilities of Sketche Engine Corpus Manager. The features of this corpus manager as software, some rules of its use, the importance of Sketche Engine for students, researchers, translators and linguists are described.

Kalit so‘zlar: *tizim, korpus menejeri, korpus, sketchengine, ilmiy-tadqiqot ishlari, dasturiy ta’minot.*

Keywords: *system, corpus manager, corpus, sketchegine, research, software.*

Zamonaviy tilshunoslikda korpus ma’lumotlaridan foydalanish tobora ommalashib bormoqda. Korpuslardan foydalanishda esa axborot texnologiyalarining o‘rni alohida ahamiyat kasb eta boshladi. Korpuslar dasturiy ta’minot shaklida bo‘lib, turli uslubda va turli dizayn asosida ishlab chiqilgan bo‘ladi. Shu sababli ham biz turli korpuslardan foydalanganimizda bizda turlicha hulosa chiqadi. Bazi korpuslar judayam yaxshi ishlab chiqilgan bo‘lsa, bazilari ustida ko‘p ishlash kerakligi ko‘rinib turadi. Korpus menejerlari ishlab chiqilgandan keyin bu kamchiliklar bartaraf etildi deyish mumkin. Korpus menejerining asosiy vazifasi, bitta dasturiy ta’minotda bir nechta korpuslarni bir vaqtda yaratib berish,

ulardan bir vaqtda foydalanish kabi imkoniyatlarib borligidir. Biz ko‘rib chiqadigan korpus menejer bu Sketch Engine korpus menejeri hisoblanib, undan deyarli butun dunyo olimlari, erkin izlanuvchilar va talabalar foydalanib kelishmoqda. Ushbu korpus menejeri yordamida mavjud korpuslar yordamida yoki yangi korpus yaratib olib, uni ustida turli xil amallar bajarish mumkin. Bu amallar xususiyatlar orqali bajariladi. Bu xususiyatlar haqida asosiy qismda batafsil to‘xtalib o‘tilgan.

Tilshunoslikda korpus (ko‘plikda *corpora*) yoki matn korpusi katta va tizimlangan matnlar to‘plamidan (hozirgi kunda, odatda, elektron saqlanadi va qayta ishlanadi) iborat til *manbayidir*. Korpus tilshunosligida ular muayyan til doirasida statistik tahlillarni amalga oshirish va gipotezani tekshirish, tildagi hodisalarni kuzatish yoki nazariy lingvistik qoidalarni tekshirish uchun foydalaniladi [2. 120-bet].

Tilshunoslikning muhim vazifalaridan biri lingvistik tadqiqotlar uchun faktik materiallar manbalarini elektron shaklda to‘plash va saqlashdir. Ayni paytda ushbu muammoni hal qilish uchun qulay tarzda saqlanadigan turli xil funksiyalarga ega maxsus texnologiyalar nafaqat katta hajmdagi matnlarni elektron shaklda saqlashga imkon beradi, balki ularni qidirish, qayta ishlash imkonini ham yaratadi. Elektron shaklda yoki korpusda matnlar to‘plamini yaratish vazifasi zamonaviy tilshunoslik uchun shunchalik muhimki, elektron matnlar amaliy tilshunoslikning maxsus bo‘limi - korpus lingvistikasining tadqiqot ob’ektiga aylanadi [3. 72-bet].

Korpuslarni qulay yaratib olish, undan foydalanish uchun uchun korpus menejeri tizimlari yaratilgan. Bunda biz tayyor korpus menejerlarida korpus yarata olamiz. Bunda bizga faqat korpus uchun kerak bo‘ladigan matnlar to‘plami kerak bo‘ladi.

Bundan tashqari korpus menejerida tayyor korpuslardan foydalanish ham mumkin. Biz bugun mavjud korpusdan foydalanish qismini ko‘rib chiqamiz. Korpus menejerining qulayliklaridan tezda bizga natija taqdim qiladi, korpus menejeri dizayni ham yaxshi ishlab chiqilgan.

Korpus menejeridan foydalanish ro‘yxatdan o‘tish orqali amalga oshiriladi. Foydalanish qismida esa asosiy xususiyatlar qismi paydo bo‘ladi.

Korpus lingvistikasining markaziy tushunchasi – lingvistik korpus – turli lingvistik parametrlar bo‘yicha belgilangan va qidiruv tizimi tomonidan taqdim etilgan maxsus tanlangan matnlar to‘plami sifatida belgilanadi. Shunday qilib, tanani quyidagicha umumlashtirish mumkin:

Korpus = matnlar + ularning belgilari.

Korpuslarni yaratishda uchun quyidagi talablar qo‘yiladi:

1) reprezentativlik (korpusdagi hodisaning chastotasi uning tabiiy tildagi chastotasiga mos kelishi kerak);

2) to‘liqlik;

3) yetarli hajm;

4) iqtisodiy samaradorlik (matnlar korpusi muammoli sohani o‘rganishda tadqiqotchining sa’y-harakatlarini tejashi kerak);

5) materialning tuzilishi;

6) kompyuterni qo‘llab-quvvatlash (matn korpusini so‘z kontekstlarini identifikatsiyalashni, statistik inventarizatsiyani, lug‘atni avtomatik qayta ishlashni va boshqalarni ta’minlaydigan ma’lumotlarni qayta ishlash dasturlari majmuasi bilan qo‘llab-quvvatlash) [4. 61-bet].

Sketche Engine korpus menejeri

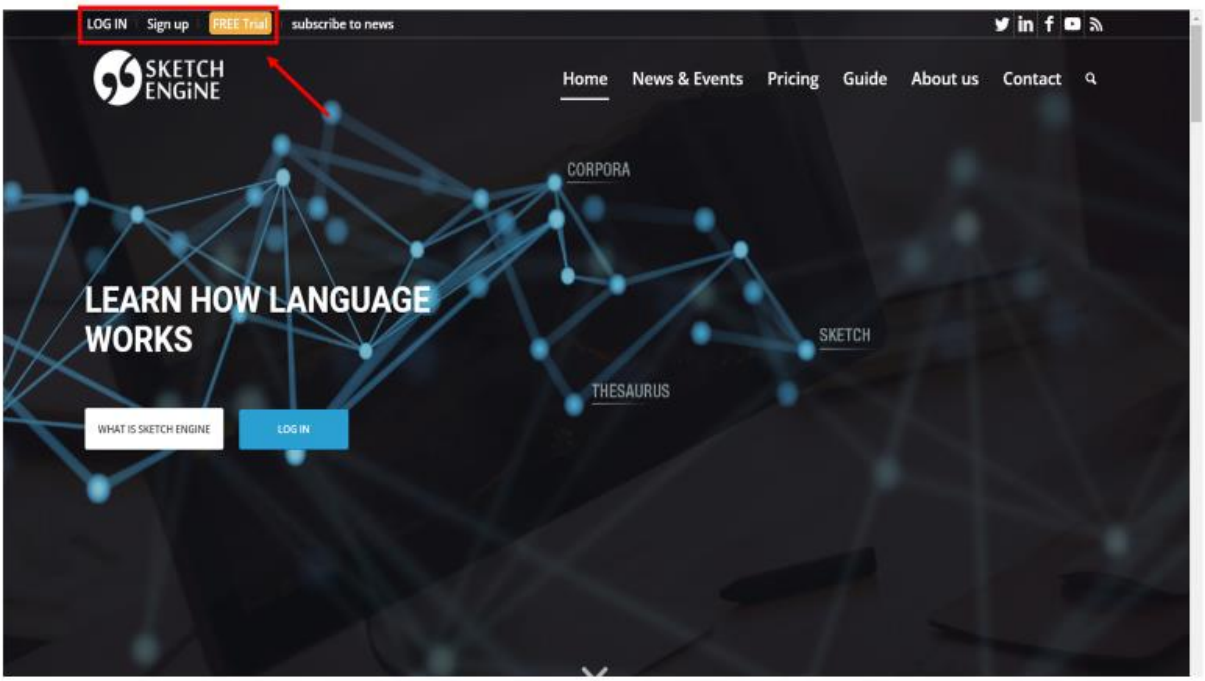
Sketch Engine korpus menejeri 2003 yildan beri Lexical Computing Limited tomonidan ishlab chiqilgan korpus menejeri va matn tahlili dasturidir. Uning maqsadi til xulq-atvorini o‘rganayotgan odamlarga (leksikograflar, korpus lingvistikasi bo‘yicha tadqiqotchilar, tarjimonlar yoki til o‘rganuvchilar) murakkab va lingvistik jihatdan katta matn to‘plamlarini qidirishga imkon berishdir.

Sketche Engine Buyuk Britaniya yoki boshqa nashriyotlar tomonidan Macmillan English Dictionary , Dictionnaires Le Robert , Oxford University Press yoki Shogakukan kabi lug‘atlarni ishlab chiqarish uchun ishlatilgan va Buyuk Britaniyaning beshta eng yirik lug‘at nashriyotidan to‘rttasi Sketche Engine-dan foydalanadi.

Ushbu korpus menejerida 90 dan ortiq tilda 600 ga yaqin korpuslar ishlab chiqilgan. Sketche Engine - bu til qanday ishlashini o‘rganish uchun eng yaxshi vosita. Uning algoritmlari tilda nima tipik ekanligini va kamdan-kam, noodatiy yoki paydo bo‘layotganligini bir zumda aniqlaydi. Korpusdagi milliardlab so‘zlardan iborat haqiqiy matnlarni tahlil qiladi.

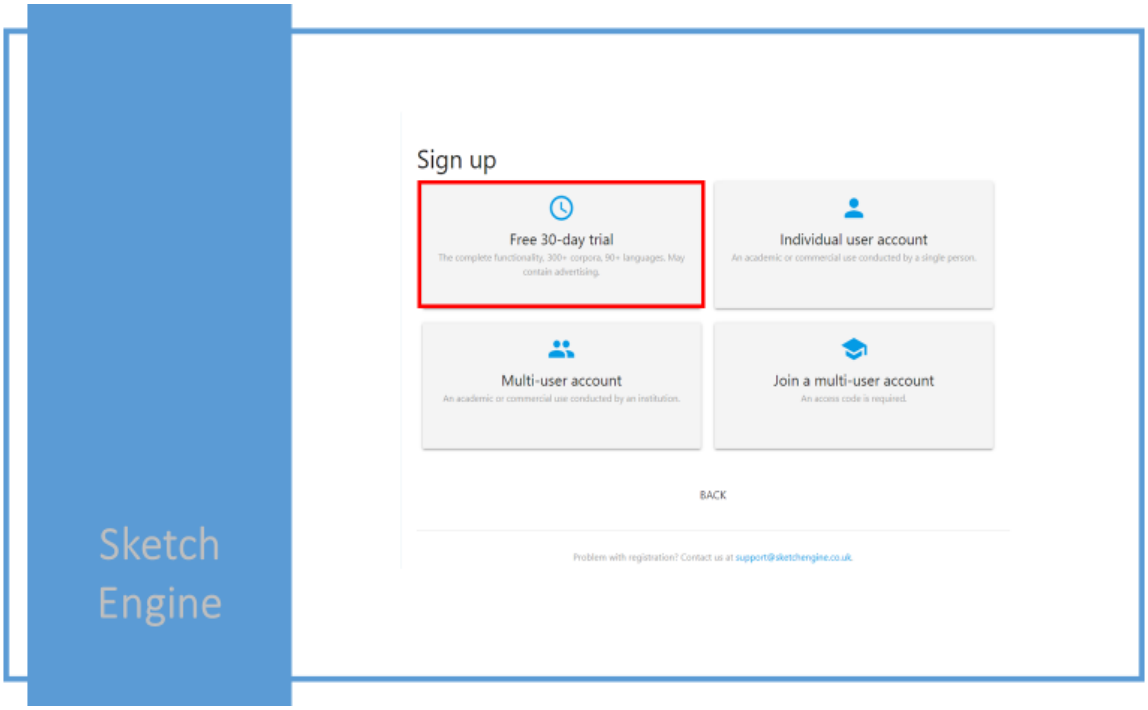
Sketche Engine avval tijoriy dasturiy ta'minot bo‘lib, keyin esa barcha xususiyatlari mavjud bo‘lgan bepul foydalanish qismi ham ishlab chiqilgan.

Quyida korpus menejeridan qanday foydalanishni ko‘rib chiqamiz:

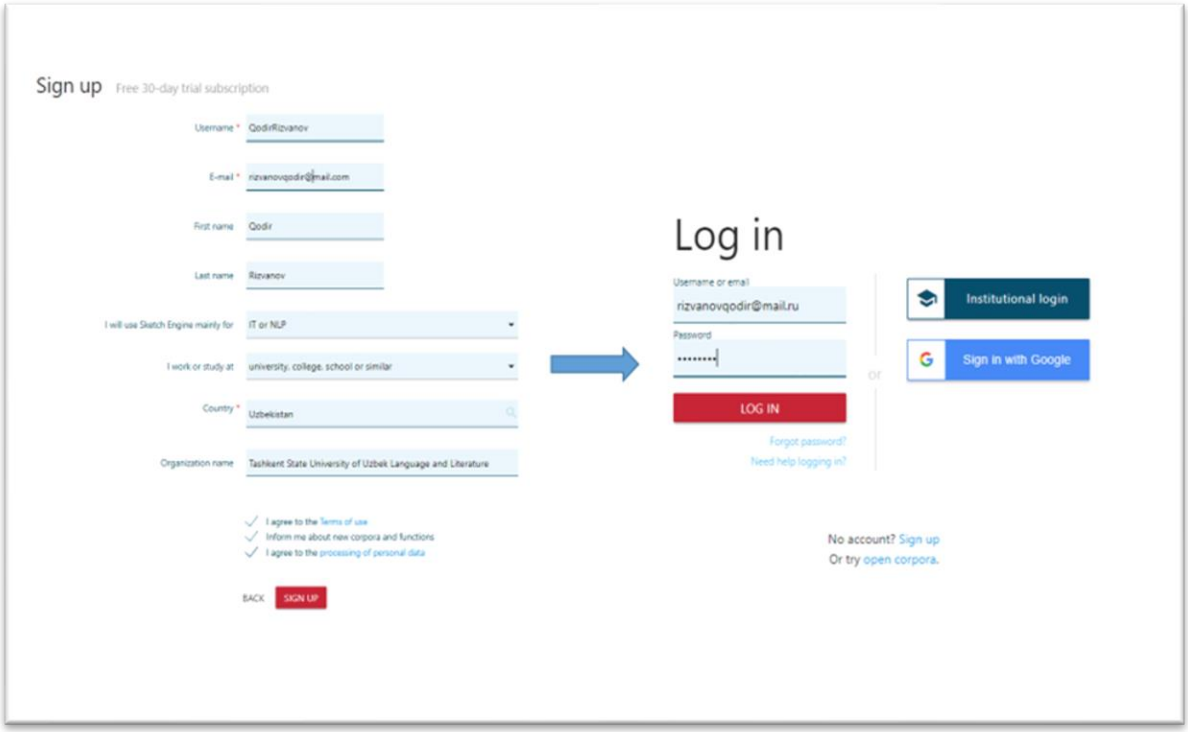


1-rasm. ***Sketch Engine korpus menejeri bosh sahifasi***

Sketche Engine korpus menejeri dasturiy ta’minotiga https://www.sketchengine.eu/ havola orqali kiriladi. Saytga kirganda eng birinchi yuqoridagi rasmdagi ko‘rinish paydo bo‘ladi. Rasmdagi belgilangan qismda biz ro‘yxatdan o‘tish yoki avval ro‘yxatdan o‘tilgan bo‘lsa, kirish qismiga o‘tiladi. Ro‘yxatdan o‘tishda 4 xil foydalanuvchi uchun harxil ketma-ketlik va harxil foydalanish qismi chiqadi. Dastlabki qism qolganlaridan farqi shundaki, qolganlari pullik hisoblanadi va imkoniyatlari ham kengroq bo‘ladi. Birinchi qism esa 30 kun davomida bepul foydalanish uchun ochiq turadi. Biz bepul versiyasini ko‘rib chiqamiz. Quyida tekin versiyasidan ro‘yxatdan o‘tishni ko‘rib o‘tamiz:



2-rasm. ***Ro‘yxatdan o‘tish***

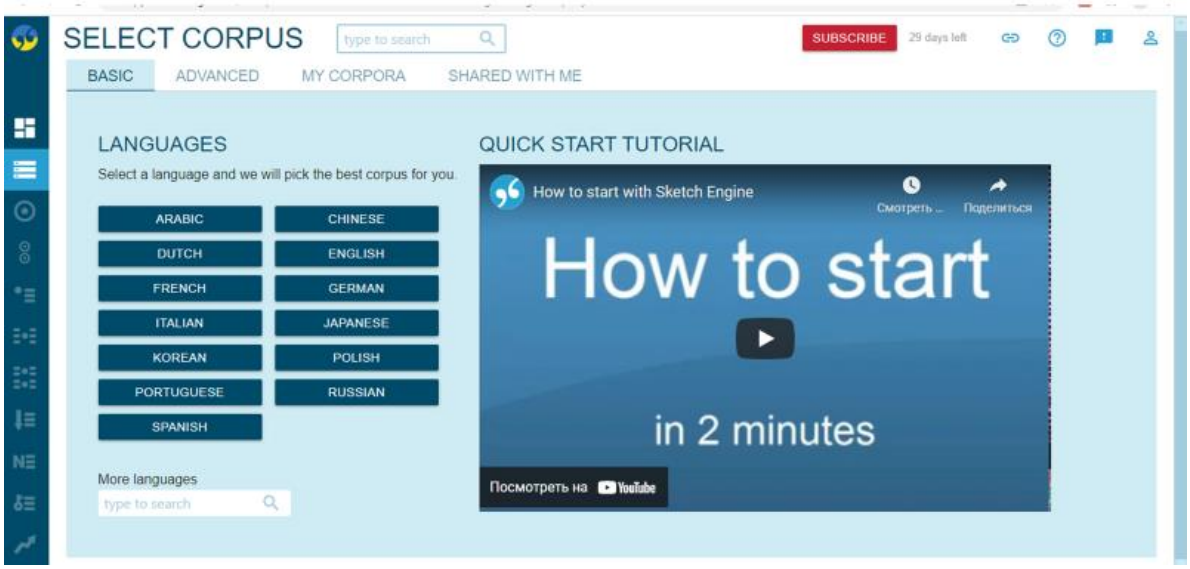


3-rasm. ***Bepul ro‘yxatdan o‘tish***

Ro‘yxatdan o‘tishda bizdan:

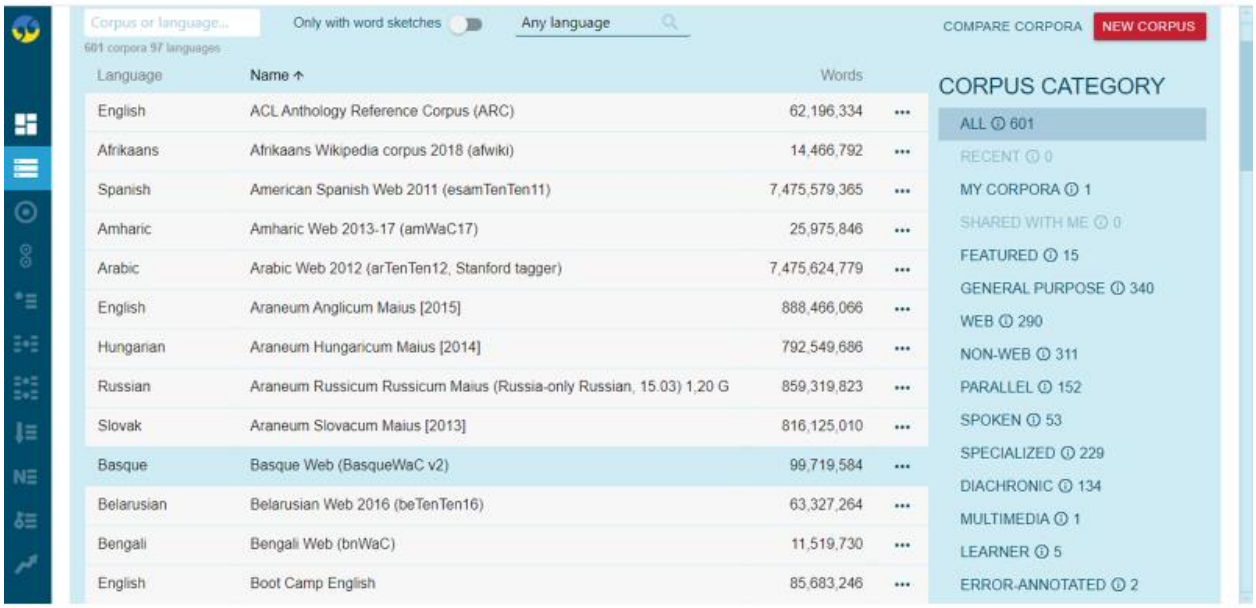
- **Username** – *har bir foydalanuvchida o‘zgacha bo‘lgan ism*
- **E-mail** – Elektron pochta
- **First name** – *Ism*
- **Last name** – *Familiya*
- *Qaysi soha vakili ekanligi*
- *Ish joyi*
- *Davlati*
- *Muassasa nomi*

so‘raladi. To‘ldirilgandan keyin tasdiqlash uchun e-mail pochtaga tasdiqlash xabari jo‘natiladi. Tasdiqlangach foydalanish uchun ishga tushadi. Quyida ko‘rib o‘tamiz:



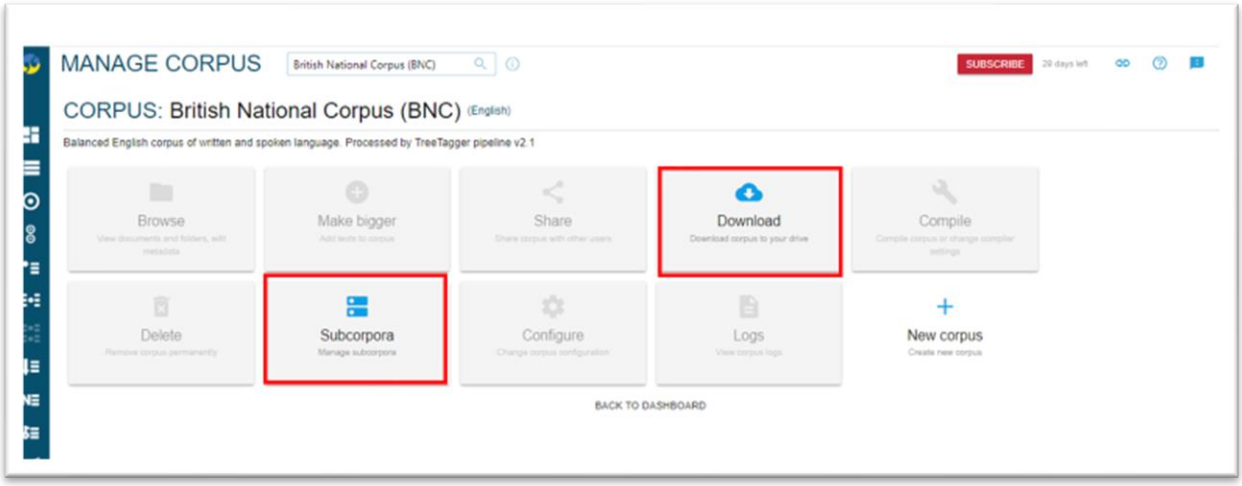
4-rasm. **Asosiy foydalanish sahifasi**

Mavjud korpuslardan birini tanlashimiz mumkin. Bunda biz kategoriya bo‘yicha ham tanlab olsak bo‘ladi. Shuningdek ko‘rishimiz mumkin, buyerda bizga korpus tili, nomi, so‘zlari ham paydo bo‘ladi.



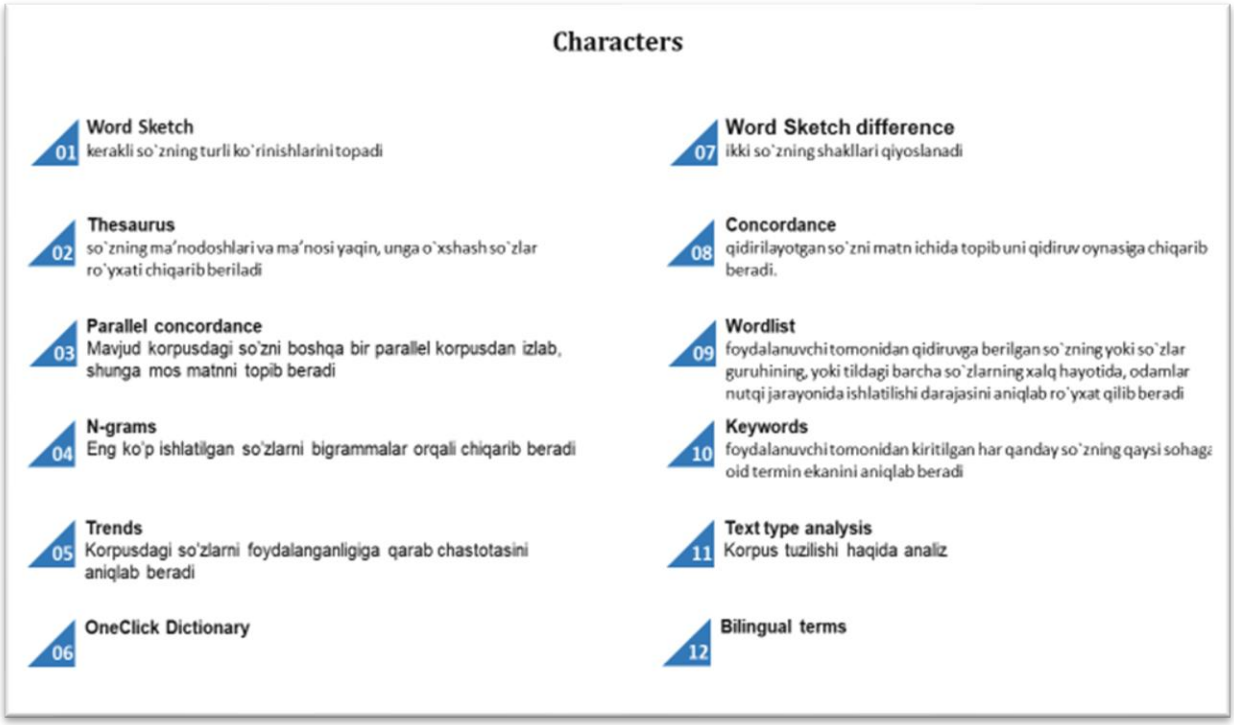
5-rasm. ***Korpuslar va kategoriyalari***

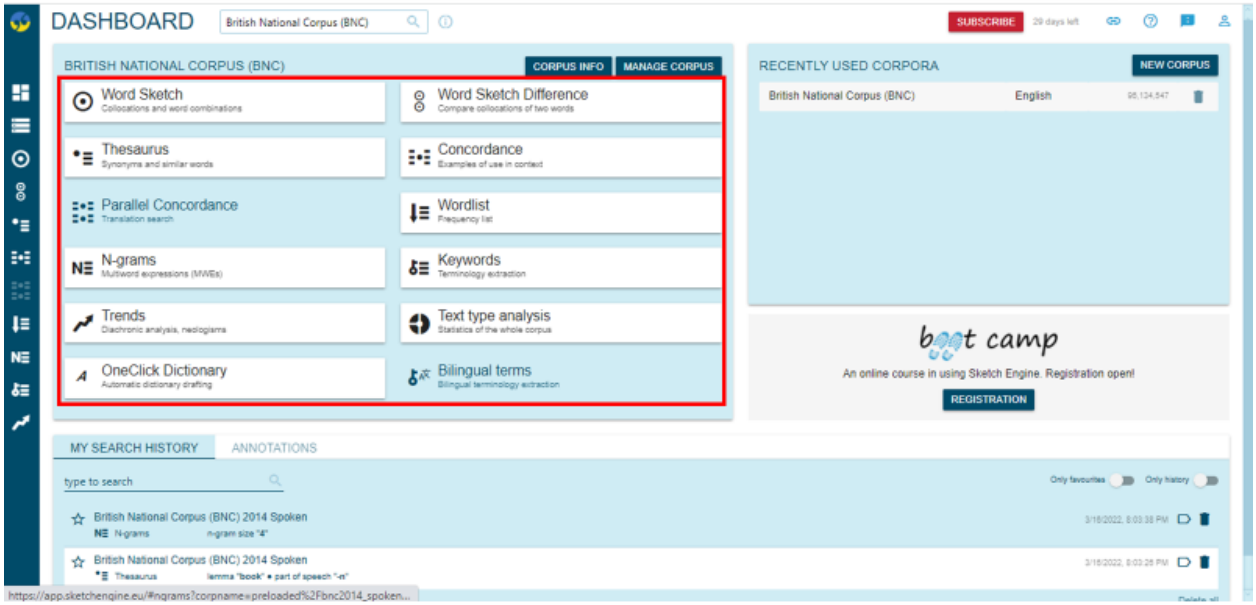
Bundan tashqari mavjud korpusni yuklab olish hamda unga subkorpus yaratish imkoniyati ham mavjud.



5-rasm. ***Korpusni yuklab olish***

Tanlagan korpusning quyidagicha xususiyatlari mavjud bo‘ladi.





Sketche Engine korpus menejeridan foydalanish hozirda ommalashib bormoqda. Bunga misol tariqasida ushbu tizimda 600 ga yaqin korpus joylashtirilganligini ham keltirsak bo‘ladi. Tilshunoslarga juda ham foydali bo‘ladigan dasturiy ta’minot. Imkoniyatlarini judayam katta qilib ishlab chiqilgan. Eng muhimi tekin foydalanish imkoniyati mavjud. Tizimning malumotlar omborida juda katta hajmdagi ma’lumotlar bo‘lishiga qaramay, tez ishlaydi. Korpus menejerida korpus yaratish va unga subkorpus yaratish imkoniyatlari esa ushbu tizim orqali foydali bo‘ladigan, dolzarb bo‘lgan korpus yaratib, uni keng ommaga taqdim etish imkoniyati ham judayam maqul keldi.

FOYDALANILGAN ADABIYOTLAR:

[1] Баранов А.Н. Введение вприкладную лингвистикую – М.: Эдуториал УРССб 2001. – С. 61.

[2] Порохницкая, Лидия Васильевна; государственный лингвистический университет. – Москва, 2004. – 195 с.

[3] Л.Ю. Щипицина информационные технологии в лингвистике С. 72-74.

[4] Гатауллин Р.Р., Гильмуллин Р.А. Контекстные правила для разрешения морфологической многозначности в корпусе татарского языка / Р.Р. Гатауллин, Р.А. Гильмуллин // Открытые семантические технологии проектирования интеллектуальных систем OSTIS-2016 – Open Semantic Technologies for Intelligent Systems: Материалы VI международной научно-технической конференции (Минск, 18-20 февраля 2016 года). – Минск: БГУИР, 2016. - С. 389-392.

[5] Невзорова О.А., Мухамедшин Д.Р., Билалов Р.Р. Корпус-менеджер для тюркских языков: основная функциональность // Труды международной конференции «Корпусная лингвистика - 2015». – СПб.: С.-Петербургский гос. Университет, филологический факультет, 2015. – С. 344-350.

[6] http://thttps://www.sketchengine.eu//

[7] https://ru.wikipedia.org/

[8] Qi Pan, A Tentative Study on the Functions and Applications of English Euphemism. – English Department, Zhenjiang Watercraft College, Zhenjiang, China: Theory and Practice in Language Studies, 2013.

[9] Rawson, Hugh. How not to say what you mean. / Hugh Rawson. – Oxford University Press, 2002. – 501 pages.