

Мова корпусних запитів – CQL (Corpus Query Language)

(На основі документації SketchEngine: [CQL - Corpus Query Language](#).)

ЗМІСТ

Синтаксис CQL

Пробіли

Атрибути (word, lemma, lc, tag) та їхнє застосування

Пошук за неповною лемою

Словоформи та леми, що починаються на.., закінчуються на.. або містять..

Пошук за неповною словоформою

Відстань між токенами, повтори

? необов'язковий токен

Українські кириличні символи в регулярних виразах

Логічні оператори в запитах

Дорівнює, не дорівнює, більше, менше

Екранування спеціальних символів: \ == !=

Використання операторів I, АБО, НЕ

Використання оператора АБО з токенами

Пошук всередині структури: containing та within

Оператор “within”

Оператор “containing”

Вкладення («гніздування») при використанні структур

NOT containing

Використання глобальних умов для порівняння токенів, пошуку однакових словоформ і лексем

Частотність з глобальними умовами

Пошук лівого і правого контексту – оператор meet

Поєднання двох запитів з оператором meet – оператор union

Структури

Початок структури

Атрибути структури

Групування

Глосарій ужитих термінів

Основи CQL

Щоб скористатися CQL, перейдіть до пошуку в корпусі та оберіть опцію CQL. У жодній іншій частині інтерфейсу CQL не працює.

Синтаксис CQL

Використовуючи CQL, можна задати комплексні критерії для пошуку одного чи кількох токенів. Критерії для кожного токена мають бути у квадратних дужках [], у форматі:

[атрибут="значення"]

Атрибут — це тип завдання для пошукової системи, а значення — його конкретна реалізація в цьому запиті.

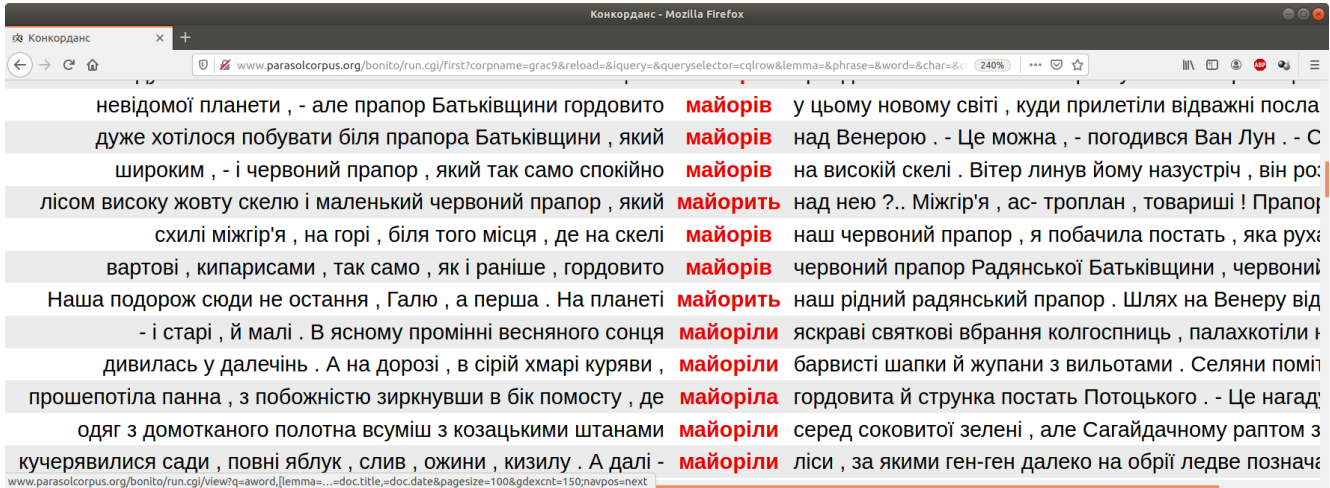
Наприклад, щоб знайти лему *бабак* (різні граматичні форми слова *бабак*), використовуйте:

[lemma="бабак"]



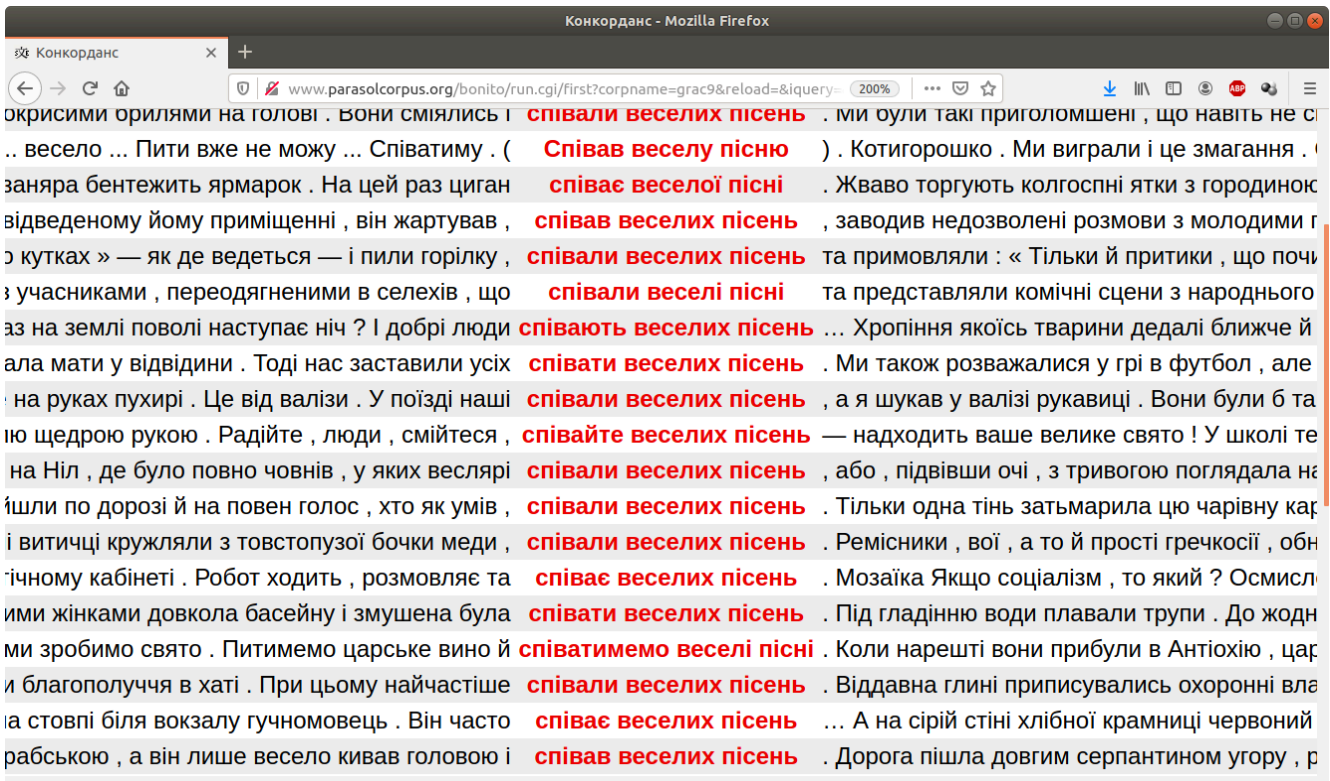
Щоб знайти лему *майоріти*, використовується такий запит:

[lemma="майоріти"]



Запит для кожного токена має міститися в окремій парі квадратних дужок. Для пошуку фрази *співати веселих пісень* з урахуванням можливих граматичних варіантів головного та залежних слів введіть:

[lemma="співати"] [lemma="веселий"] [lemma="пісня"]



Пробіли

Пробіли не виконують жодної функції в SQL. Можна використовувати пробіли, щоб зробити текст зручнішим для читання. Такий запит нічим не відрізняється від попереднього:

```
[ lemma="співати" ][ lemma="веселий" ][ lemma="пісня" ]
```

Однак вираз не спрацює, якщо в ньому буде декілька пробілів поспіль.

У лапках не повинно бути жодних пробілів. Наприклад, такий пошук не дасть результатів, оскільки лема ніколи не може починатися з пробілу:

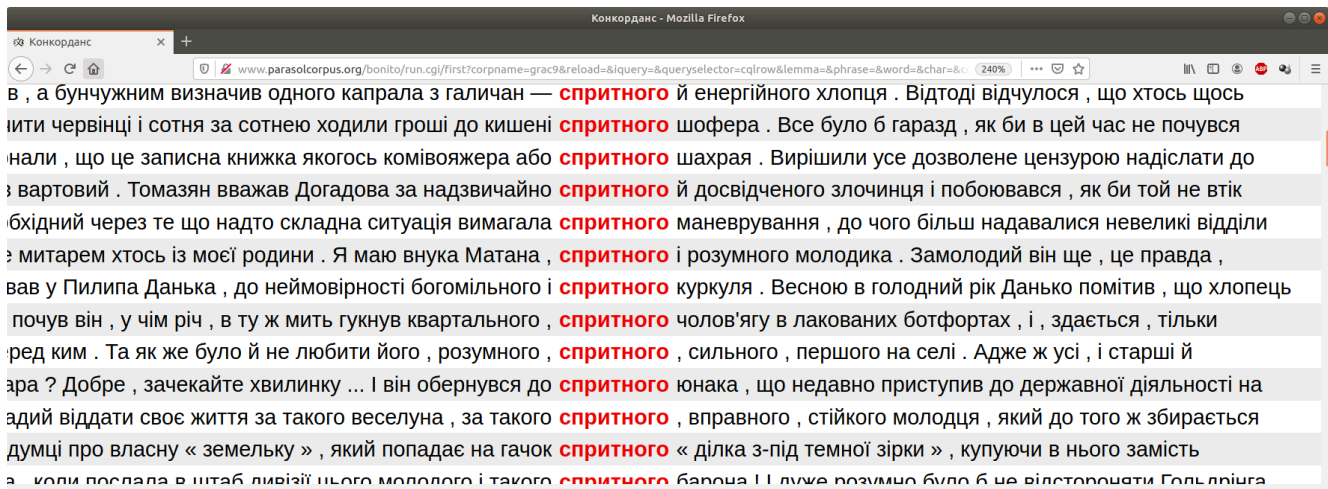
```
[lemma=" хто"]
```

Атрибути (*word*, *lemma*, *lc*, *tag*) та їхнє застосування

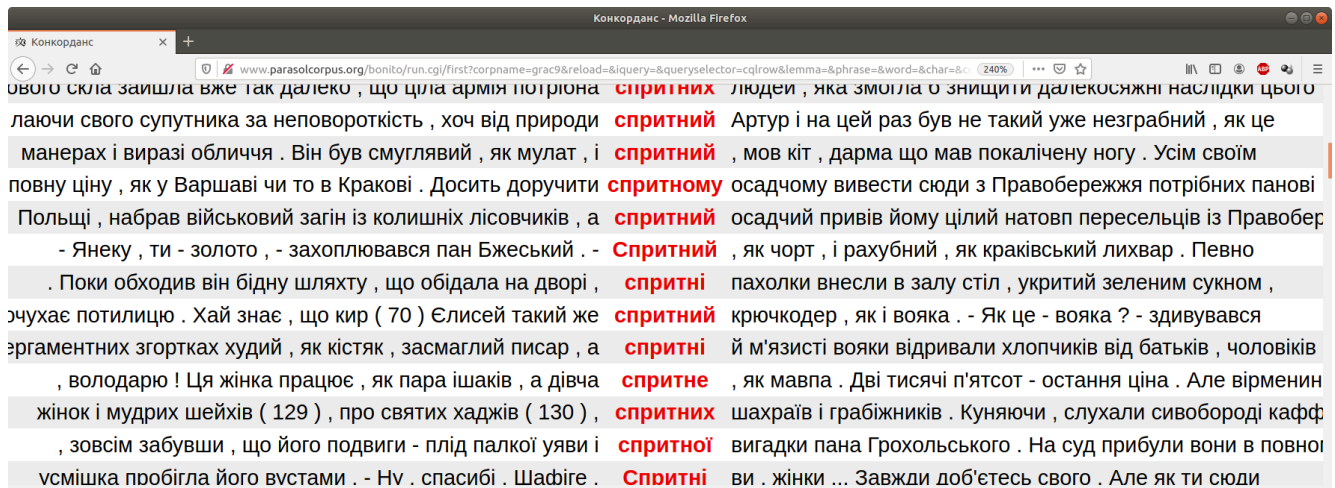
У таблиці розглянуто найпоширеніші атрибути: *word*, *lemma*, *lc* (*lower case*) і *tag*, їхнє призначення та приклади оформлення в запитах.

ЗАВДАННЯ	SQL КОД	РЕЗУЛЬТАТ
Знайти приклади зі словоформою <i>спритного</i> (тільки з малої літери)	[word="спритного"]	Конкорданс словоформи <i>спритного</i> , з малої літери
Знайти приклади всіх форм лексеми <i>спритний</i>	[lemma="спритний"]	Конкорданс усіх словоформ лексеми <i>спритний</i> : <i>спритного</i> , <i>спритному</i> , <i>спритна</i> , <i>спритні</i> тощо
Знайти приклади зі словоформою <i>спритний</i> малими літерами та з великої літери на початку речення	[lc="спритний"]	Конкорданс словоформи <i>спритний</i> з малої літери та з великої літери на початку речення (приклади з прізвищем <i>Спритний</i> в середині речення не потраплять до вибірки)
Знайти приклади всіх словоформ з тегом <i>prop</i> (власні назви)	[tag=".*prop.*"]	конкорданс різноманітних слів з тегом, що містить <i>prop</i> : <i>Ганна</i> , <i>Петрівна</i> , <i>Короленко</i> , <i>Київ</i> , <i>Десна</i> тощо
Знайти приклади всіх словосполучень прикметників на позначення кольорів з іменником	[semtag=".*color.*"][tag="noun.*"]	

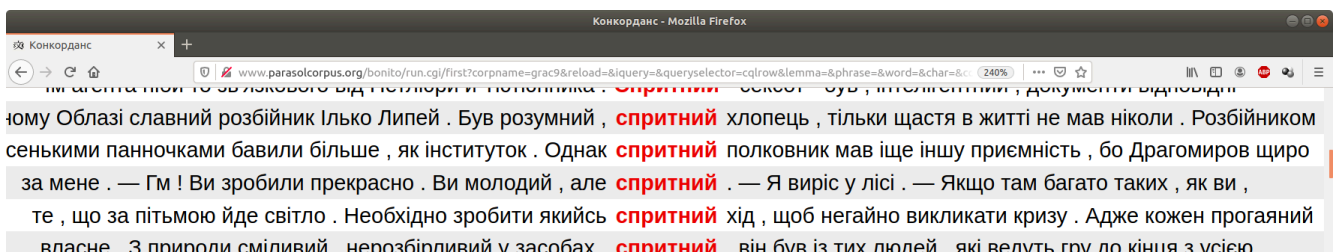
[word="спритного"]



[lemma="спритний"]



[lc="спритний"]



Починаючи з 13 версії до корпусу ГРАК додана семантична розмітка, доступна за атрибутом *semtag*, завдяки якій можна знаходити токени тих чи інших семантичних категорій (докладніше у [статті В.Старка](#)). Приміром, за цим запитом можна знайти усталені порівняння, в яких компарандумом є назви тварин, а компаратумом є будь-який прикметник:

```
[tag="adj.*"&tag!=".*pron.*"] [lemma="як|наче|мов|ніби"] [semtag=".*animal.*"]
```

	Лемма	Частота	Відносні ?
1 <input type="checkbox"/>	голодний як вовк	89	0,05 ...
2 <input type="checkbox"/>	відомий як представник	72	0,04 ...
3 <input type="checkbox"/>	червоний як рак	70	0,04 ...
4 <input type="checkbox"/>	худий як тріска	54	0,03 ...
5 <input type="checkbox"/>	більше більший як сорок сорока	41	0,02 ...
6 <input type="checkbox"/>	здоровий як бик	33	0,02 ...
7 <input type="checkbox"/>	сивий як голуб голубити	25	0,01 ...
8 <input type="checkbox"/>	червоний мов мова рак	23	0,01 ...
9 <input type="checkbox"/>	німий як риба	21	0,01 ...
10 <input type="checkbox"/>	хитрий як лис	17	< 0.01 ...
11 <input type="checkbox"/>	сивий як лунь	16	< 0.01 ...
12 <input type="checkbox"/>	голодний як пес	15	< 0.01 ...
13 <input type="checkbox"/>	задоволений як слон	14	< 0.01 ...
14 <input type="checkbox"/>	голодний як собака	13	< 0.01 ...
15 <input type="checkbox"/>	здоровий як бугай	12	< 0.01 ...
16 <input type="checkbox"/>	менше менший як сорок сорока	12	< 0.01 ...
17 <input type="checkbox"/>	злий злити як собака	11	< 0.01 ...
18 <input type="checkbox"/>	спокійний як удав удати	10	< 0.01 ...
19 <input type="checkbox"/>	спокійний як слон	9	< 0.01 ...

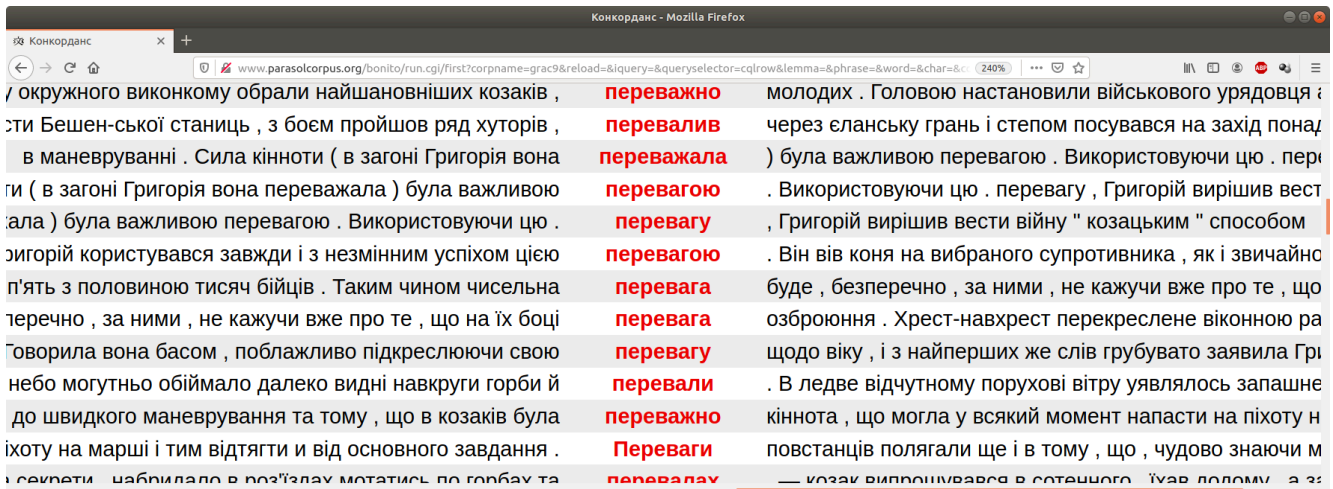
Пошук за неповною лемою

Словоформи та леми, що починаються на..., закінчуються на.. або містять..

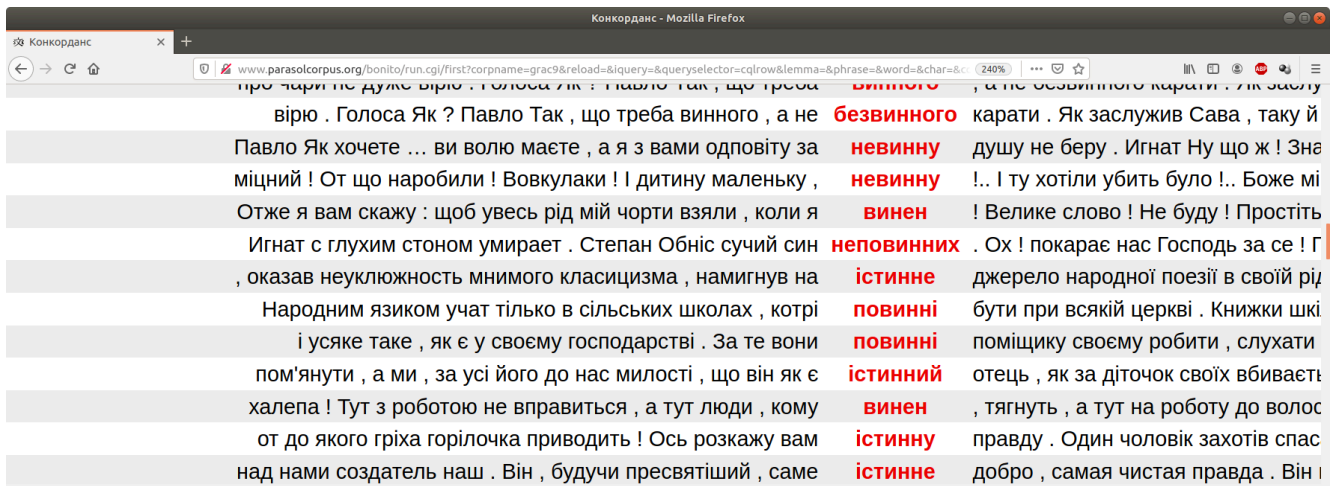
У лапках можна застосовувати [регулярні вирази](#).

ЗАВДАННЯ	CQL КОД
Леми, що починаються на <i>перева-</i>	[lemma="перева.*"]
Леми, що закінчуються на <i>-инний</i>	[lemma="*.инний"]
Леми з трьох літер, які починаються на п- і закінчуються на -к	[lemma="п.к"]

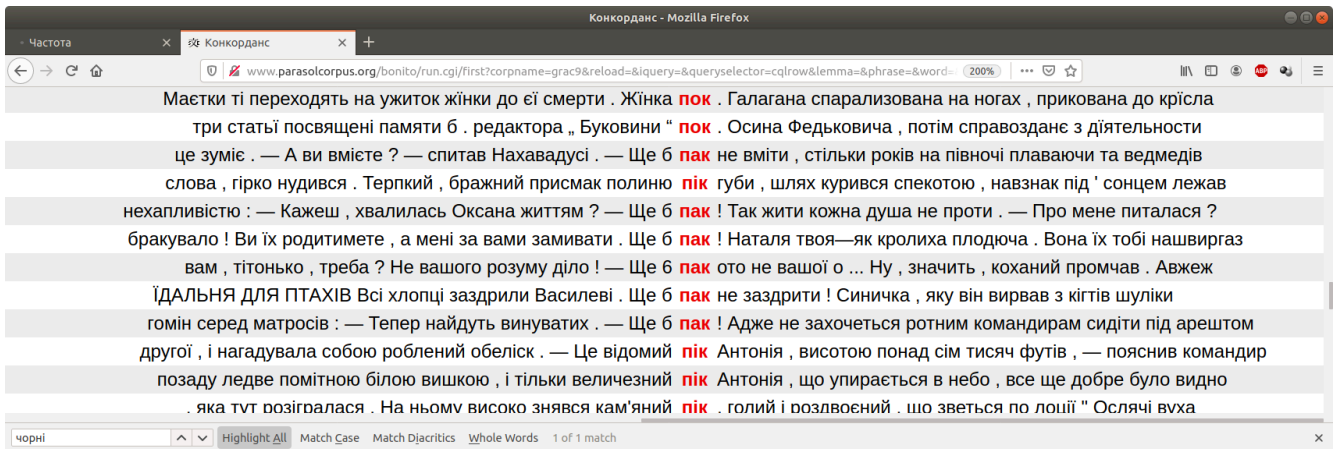
[lemma="перева.*"]



[lemma="*.инний"]



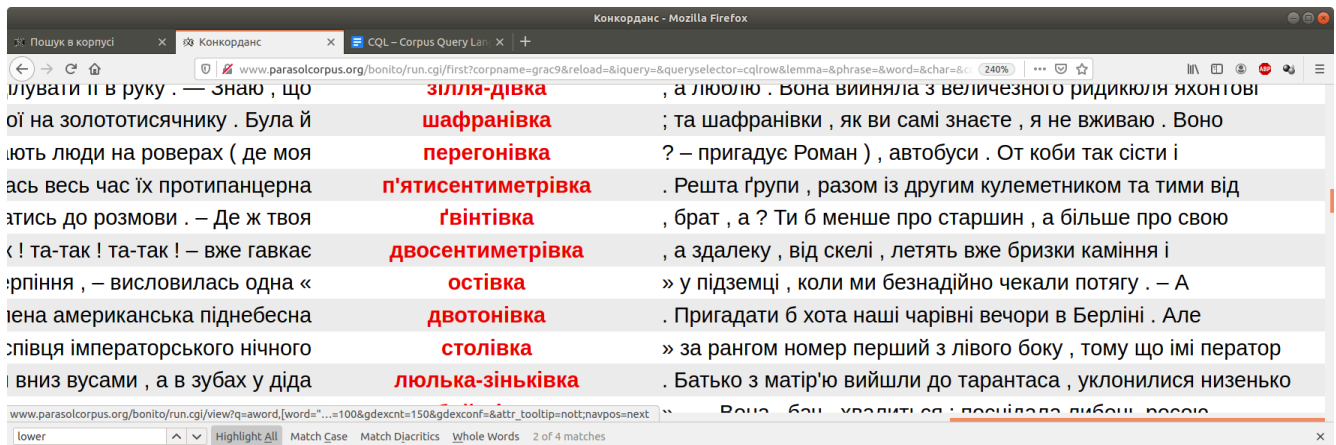
[lemma="п.к"]



Для пошуку словоформ, що написані всіма малими літерами, можна скористатися додатковою умовою всередині токена – `[:lower:]`. Додати умову до літер можна так:

`[word="[:lower:].*івка"]`

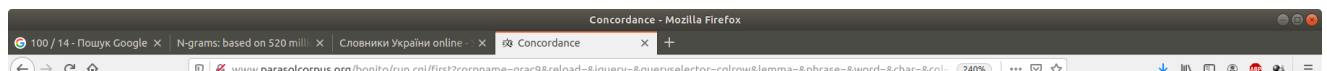
Цей запит виконує пошук усіх словоформ, що закінчуються на `-івка` і написані всіма малими літерами:



Пошук за неповною словоформою

Пошук за словоформою потрібен, коли користувача цікавить лише певна граматична форма якоїсь лексеми. Для швидкої перевірки вживаності певної форми, наприклад, наказового способу першої особи множини дієслова спати "спімо":

`[word="спімо"]`



Такі результати можна отримати й за допомогою простого запиту. Однак, якщо виникає потреба скористатися регулярним виразом, запит необхідно писати в полі CQL. Так, наступним виразом можна скористатися, щоб знайти консонантні рими до словоформи “молитва”:

[word="*.итва"]

жар , щоб не розгорілось вогнище . Ось що пише у статті " **Битва** за Україну " (бюлетень " Гірка правда ") С. Хмара :
 років тому війська Сулеймана-паші увійшли у Візантію . **Битва** точилася цілий день і надвечір , коли ворог уже виснажився
 ніяк не зменшилась . Я обстриг свої патли , а ножиці та **бритва** в руках Інги надали моїй новій зачісці якогось панківського
 молодші . Емісар твоїх років — чого б інакше була така **гонитва** за тобою ? До того ж , серце моє у типовому положенні
 Osm . Spr . 49 ; Adalb . Jezuk 48 . 56 . У нього язик як **бритва** острій . (Наг .) Говорять про дотебного з'їдливого чоловіка
) Пор . т . І , Бог 148 . Даль І , 297 . 191 . За Богом **молитва** , за царем служба не пропаде . Пор . Даль І , 297 . 192
 грішному тай мельничку горішному . (Тухля) . Жартлива **молитва** . 402 . Божи , Божи , злізь , подивись , тай назад вернись
 Даль І , 83 . 11 . Богатого і серп голит , а бідного і **бритва** нехоче . Пор . Даль І , 40 . 25 . Добре багатому красти
 Бристи 1 . Брив , брив , при березі утопивсі . (Жидач .) **Бритва** 2 . Бритва й жінки не позичай нікому . (Снят .) Правило
 Брив , брив , при березі утопивсі . (Жидач .) Бритва 2 . **Бритва** й жінки не позичай нікому . (Снят .) Правило життєвої
 Моленіє 1 . Яке моленіє , таке спасеніє . (Убинє) . Нецира **молитва** не причиняється до спасення душі . Молебен 1 . Дав бим

Якщо потрібно знайти всі словоформи певної лексики, звичайно ж, доречно використовувати атрибут *lemma*, однак для лексем, що їх немає в словнику ВЕСУМ <https://r2u.org.ua/vesum/>, за яким розмічений ГРАК, пошук за лемою не дасть результатів. Зокрема, відсутні в словнику помилкові вживання. Приміром, для вибірки помилково вживаної лексики “торгівельний” пошук доцільно проводити так:

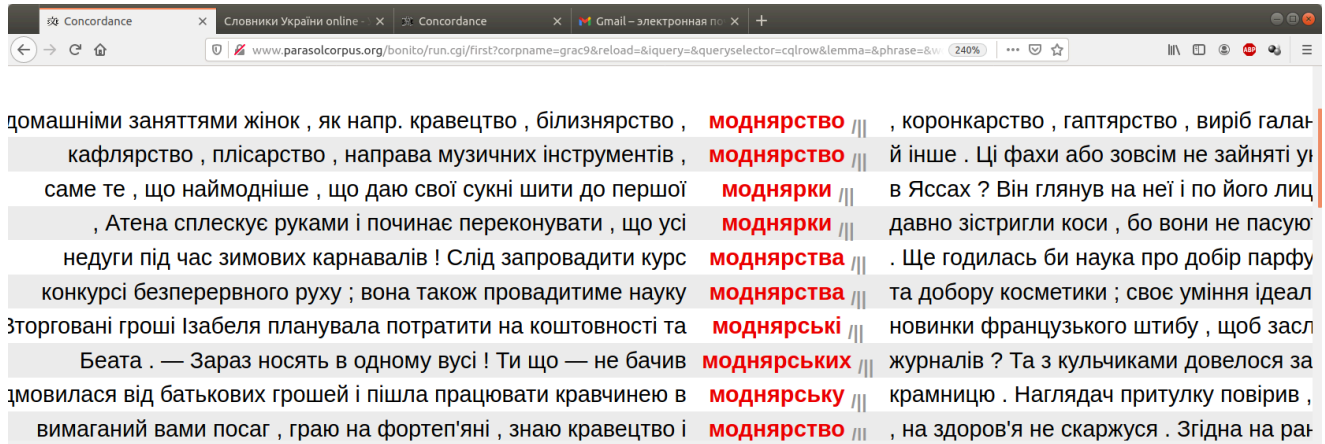
[word="торгівельн.*"]

:вітами розводять руками чим торгувати . Спекулянти без **торгівельних** технікумів та інститутів , хіба що з швидко наї
 тишнього начальника управління торгівлі , нині директора **торгівельно-комерційного** підприємства Валентину ПИЛИПОВЕЦЬ та го
 олівця , бо , крім школярів , не потрібно забувати нашим **торгівельникам** , у місті - три училища , де навчаються сотні .
 . . . як у нас . Чому адресує це питання в першу чергу **торгівельникам** , владним структурам . Кульковий заряд до ру
 ого і районних центрів , колгоспна верхівка годувалася на **торгівельних** базах та з чорних ходів магазинів ? Перебира
 пективи відносин Росії і НАТО , вступ Росії до Всесвітньої **торгівельної** організації тощо . У планах Путіна - зустріч з г
 світових економічних організацій і форумів - Всесвітньої **торгівельної** організації , Вишеградської групи тощо , в яки
 аївському , Дрогобицькому , Перемишлянському районах **торгівельна** надбавка виходить далеко за межі 15% . А інс
 кавав : - Чого дивуватися ? Можливо , бабуся займається **торгівельним** бізнесом , а телефоном керує рухом товарів .
 дуже помпезно . На той час похід за покупками до цього **торгівельного** дому був для львів'ян приємністю та чимось ф
 ласична система . Є капітал основний - земля . Є капітал **торгівельний** , який став фінансовим , обіговим . З'явилася

У газетних текстах ГРАК-9 виявлено 167 таких помилкових вживань.

За лемою неможливо знайти деякі архаїзми, неологізми, okazionalizmi тощо, яких немає в словнику. Також у корпусі не лематизовано ненормативні, обценні лексеми, слова з друкарськими та правописними помилками. Нижче наводяться приклади запитів для пошуку за словоформою саме таких раритетних лексем, що не внесені до словника:

[word="модняр.*"]



Атрибут *lemma* у цих tokenів є порожнім

Відстань між токенами, повтори

Квадратні дужки [] означають 'будь-який токен'. Фігурні дужки { } слугують для позначення повтору попереднього токена.

Завдання	SQL код	Результат
Знайти приклади словосполучення з лексем <i>співати</i> та <i>пісня</i> , між якими міститься один токен	[lemma="співати"] [] [lemma="пісня"]	<i>Співати люблю пісень співав інколи пісень співати хочеться пісень</i>
Знайти приклади словосполучень лексем <i>мати</i> та <i>думка</i> на відстані від 2 до 4 tokenів	[lemma="мати"] [] {2,4} [lemma="думка"]	<i>Має свою власну думку має час зважувати свої думки</i>
Знайти приклади словосполучень з лексем <i>пити</i> та <i>вода</i> на відстані рівно двох tokenів між ними, що є прикметниками	[lemma="пити"] [tag="adj.*"] {2} [lemma="вода"]	<i>пити достатньо чистої води пив холодну газовану воду п'є забагато газованої води</i>
Знайти приклади словосполучень з лексем <i>пити</i> та <i>вода</i> на відстані двох або більше tokenів між ними, що є прикметниками	[lemma="пити"] [tag="adj.*"] {2,} [lemma="вода"]	<i>п'ю цю гірку небесну воду ніколи, здається не пив такої</i>

смачної холодної
лісової води

Пошук на відстані двох слів:

[lemma="пити"] [tag="adj.*"]{2}[lemma="вода"]

юго каламутна і дуже солодка до пиття , і не обридн	пити ту святу воду	, і не хворіють від неї , і не шкодить в
йманих висипах великої ріки нашого народу , обидва	пили її м'яку воду	. Він щасливий від того , що я пливу
а половина червня з великою спекою . Вася без кінця	пив холодну содову воду	і був захворів під кінець , але хозяйк
ико бував тут не раз і добре пам'ятав криницю , з якої	пив таку смачну воду	. Хутко перейшов на другий бік і пішс
кажучи , сер , — відповів я , — ні того , ні другого ,	п'ю чисту холодну воду	. — О-о-о , чоловіче ! — вигукнув Дж
гну рекламу , що я в моєму життї (ще й до сьогодні)	пив ріжні мінеральні води	, але ці кавказькі вважаю найбільш п
чні покидьки дихання . Протягом тижнів , Лоттхен , ми	п'ємо застоюну теплу воду	, і то в обмеженій кількості . Її доведу
з , що можна дихати , жадібно ковтав повітря , неначе	пив холодну газовану воду	, й слухав чарівну пісню . А голос то :
ечка . Я витяг відро води . Ніколи мені не доводилося	пити такої смачної води	! Я пив і пив . Потім налив у пляшку і
уть , бібліотека покійного лорда Гленмора . — Ви вже	пили цю їхню воду	святого Нініана , шановні панни ? —
уважав у таку погоду ! Їй-богу , я б , бувши вами , не	пив більше цієї води	. Я ВИПИВ її один лише ковток , і то

Пошук на відстані до двох слів:

[lemma="пити"] [tag="adj.*"]{0,2}[lemma="вода"]

ив натщесерце , ковтаючи дим , до рво-ти , а потім	пив воду	, міцно настояну на тютюні . Це робилося день у ден
З тропіках , знемагаючи від спеки , він безперервно	пив воду	і крижем лежав біля машинних люків під рукавом па
ж і , підобідуючи втретє , згадав , що вона й досі не	пила води	. І вони поїхали ще до водоперегінної будки . Там , б
ю стояли тикви , там були дві жниці . Вони жадібно	пили воду	, заломивши голови й перехиливши тикви до рота . Т
і . Знесилені мордуванням , майстри важко сопли й	пили холодну воду	, витираючи спітнілі обличчя . - Лягай , стерво , - нак
ним потом . Він борсався в ліжку , як божевільний ,	пив холодну воду	і , нарешті , вигдавав чудовий вихід : « Викличу Грохо
ковтнув думку : пард ! Десь там крикнула каня , що	п'є воду	тільки з неба , день закурився , збуранів ... З отрутн
почувши гучний голос свого пана : — Сину мій , не	пий води	; не пий її , Санчо , вона вб'є тебе . Дивись , ось у
: , складається з людей різних націй — тут такі , що	п'ють ніжну воду	славетного Ксанта ; горяни , що толочать масильськ
ии , лісами , ланами , співаючи тут , зітхаючи там , і	питимемо кришталеву воду	з криниць , чи з прозорих струмків , чи з повноводих
здавив . Мучила мене спрага , страх хотілось мені	пити води	, але її поблизу не було . Встав і почав навкруги шук

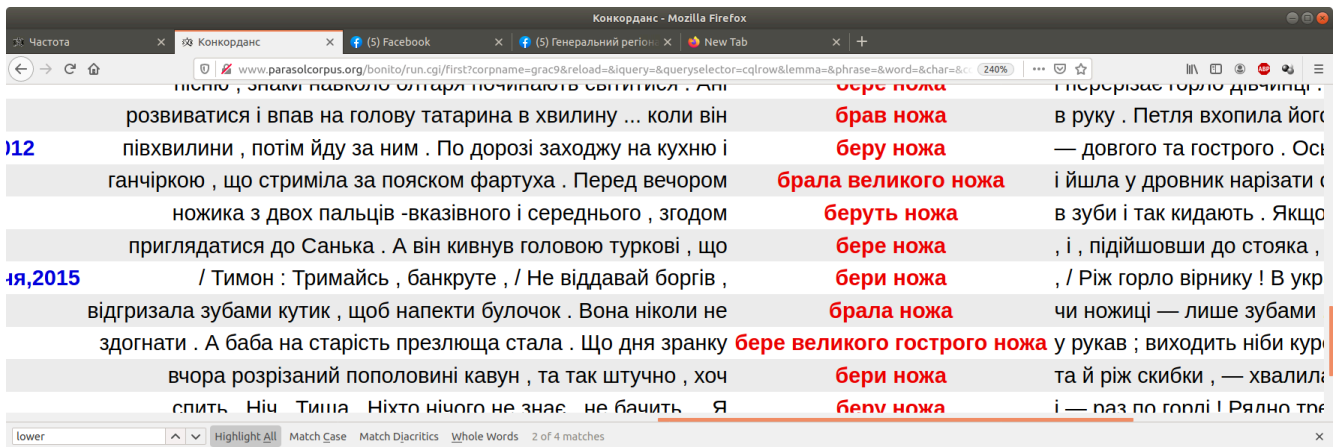
[lemma="пити"] [tag="adj.*"]{2,}[lemma="вода"]

парасолію . Також варто проводити нейтралізацію содою , пити содові мінеральні негазовані води , наприклад боржомі , п		
приймати душ навіть кілька разів на день . Слід продовжувати пити негазовані содові мінеральні води . Раджу також , перш ні		
процедуру двічі на день - через півгодини після обіду і наніч . Пити більше мінеральної води - віруси , що викликают		
дівчатка та хлопчики замість молока та молочних продуктів п'ють солодкі кольорові води . У меню дітей повинно		
двічі на день - через півгодини після обіду і на ніч . 4 . Пити більше мінеральної води - віруси , що викликают		
мого не куштував такого солодкого хліба та іншої їжі , не пив такої солодкої води , як була на тій трапезі		
Мальті вода несмачна . Зовсім не для пиття . Здається , ніби п'єш розведену морську воду . І хліб пересолений . Н		
п'єте чисту воду ? І перестати хворіти , старіти – почати пити чисту живу воду , завдяки домашній міні-		
Богоявлення . Цього дня родини знову сідають до Святої вечері і п'ють богоявленську свячену воду , яку освячують звечора		

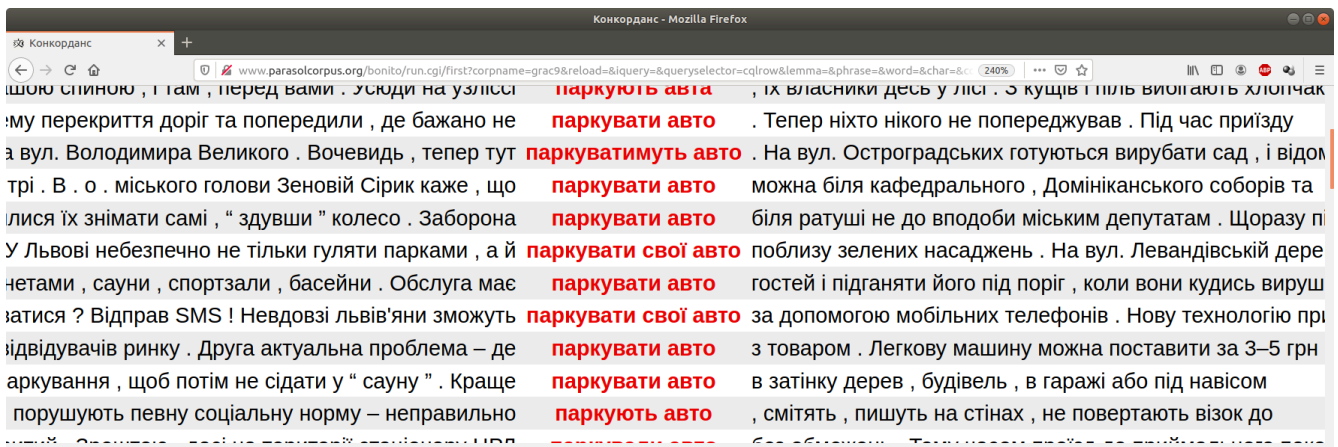
Будь-яка кількість повторів токена (0 або більше) позначається зірочкою * після нього.

Пошук на відстані будь-якої кількості прикметників:

[lemma="брати"] [tag="adj.*"]* [word="ножа"]



? необов'язковий токен



Токен можна помітити як опційний за допомогою знаку питання ? після квадратних дужок.

Завдання	CQL код
Знайти приклади словосполучення <i>паркувати авто</i> або <i>паркувати своє авто</i>	<code>[lemma="паркувати"][lemma="свій"? [lemma="авто"]</code>
Альтернативний спосіб пошуку без знаку питання	<code>[lemma="паркувати"][lemma="свій">{0,1} [lemma="авто"]</code>

`[lemma="паркувати"][lemma="свій"? [lemma="авто"]`

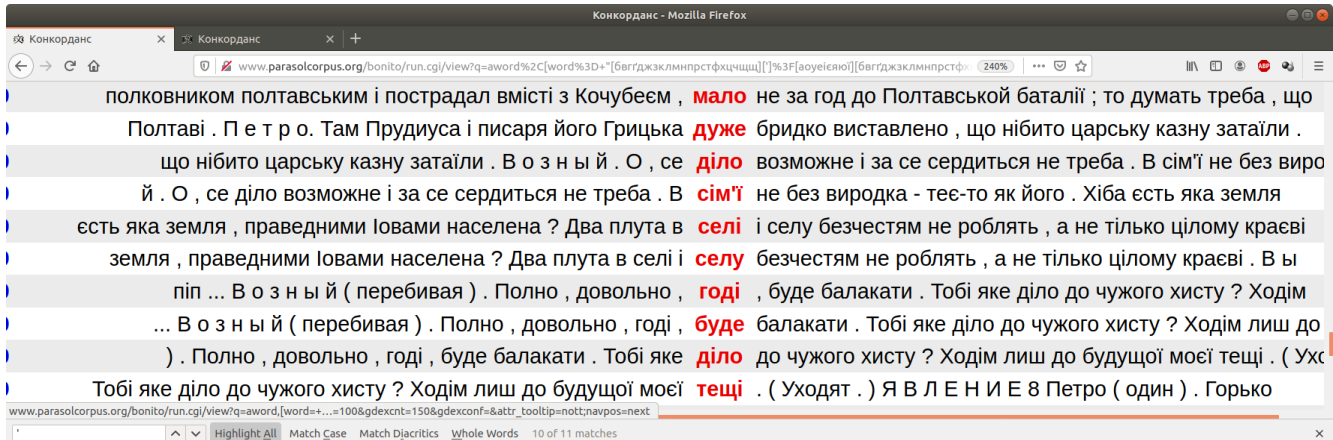
або

`[lemma="паркувати"][lemma="свій">{0,1} [lemma="авто"]`

Знайти всі словоформи за шаблоном:

приголосний + *апостроф* (опційно) + *голосний* + *приголосний* + *апостроф* (опційно) + *голосний*

[word=



"[бвггджзклмнпрстфхцщ]]?[аоуеієяюі][бвггджзклмнпрстфхцщ]]?[аоуеієяюі]"

Українські кириличні символи в регулярних виразах

Під час опису шаблону українських слів за допомогою регулярних виразів, слід враховувати специфіку української кирилиці. Перерахування символів від *а* до *я* – [а-я] – охопить не всі українські літери, натомість усі українські малі літери описує вираз: [а-яіієґ-]*.

Усі українські словоформи з урахуванням великих і малих літер можна охопити таким запитом:

[word="[А-ЯІІЄҐа-яіієґ-]*"]

або

[word="(?)[а-яіієґ-]*"]

Логічні оператори в запитах

Дорівнює, не дорівнює, більше, менше

CQL підтримує такі оператори порівняння:

дорівнює	=
менше або дорівнює	<=
більше або дорівнює	>=

не дорівнює	!=
не менше або дорівнює	!<=
не більше або дорівнює	!>=
дорівнює буквально	==
не дорівнює буквально *)	!==

*) Подвійний знак рівняння означає, що вираз у лапках має сприйматися не як регулярний, а як звичайний текст.

Екранування спеціальних символів: \ == !=

Деякі символи (крапка, зірочка, знак питання, слеш, плюс, лапки) мають спеціальні функції в регулярних виразах. Якщо вираз у лапках має сприйматися не як регулярний, а як звичайний текст, його необхідно екранувати подвійним знаком рівняння або слешем:

CQL	Результати
[word="."]	однолітерні слова
[word=="."]	Крапки
[word="\."]	Крапки

Зауважте, що в разі використання

==

!==

усе ж необхідно екранувати два символи: лапки (") та зворотний слеш (\).

Запит для пошуку слів, які закінчуються на *-ar* і яких немає в словнику ВЕСУМ

[word=".*ar"&tag="unknown"]

The screenshot shows a search results page with the following entries:

Text snippet	Word	Description
працівники кримінальної поліції Маріуполя , які самі збували букву „ б “ тріпав . Знімав квартиру в Києві і тягав туди	наркотовар	. До операції із затримання зловмисників бу
Городенку чи Коломию , або пошукати попутку на сайті «	шмар	, поки вона Сонечку носила в животі . Всі це
мудехар Одрозу навпроти собору – червоні ворота у севільський	бла-бла-кар	» . Добиратись , якщо автобусом – десь год
, як у касі , але буває дорожче – десь на одне євро . У	алькасар	. І яке щастя , що квитки заздалегідь купили
алькасар онлайн-квиток коштував 12 з половиною євро . Словом «	алькасар	онлайн-квиток коштував 12 з половиною євр
замків , збудованих за правління маврів . Але севільський	алькасар	» (з арабської « фортеця ») називають бага
музеї . Палац Насрідів – місце , де забувається Севільський	алькасар	будували християни для короля Кастилії Пе
мовляв , усе село горювало , що така гарна й молода такий «	алькасар	. Тут розписи з орнаментів і арабського пись
посприяти у поверненні коштів із відповідного рахунку .	тігар	» собі взяла . « А ми дуже добре жили , – ро
Біля озера спускалися неправильно , бо там був сніговий	хабаріхабар	Деталі дивіться у сюжеті каналу 402 . На Фр
	кулуар	, не можна було йти . Але він нас туди повів

Запит для пошуку всіх нелематизованих словоформ: [tag="unknown"&word="[А-ЯіЄҐа-яїєґ-]*"]

Результати застосування такого запиту в підкорпусі текстів Анатолія Перепаді (спроба пошуку авторських оказіоналізмів):

голосу Плінія Старшого (чи Молодшого ?) , коли перед очима	пломенилася	лава , а на голову сипався попіл і пікірували вулканічні
бюрко і два стільці . Лежачи , Микола Олексійович брав з	однієї	, знайомої мені ще на Коцюбинського , газету або журнал
гідадього Дон Кіхот з Ламанчі » ... Ось уже з каретки моєї «	оптимі	» виповзає : « Кінець першій частині » ... А далі , далі
сторінок обсягу . Та й посередині її зяє , як у щелепі	ламанця	від градового бою , бреш . Переклад розділів XXI-XXXV
посередині її зяє , як у щелепі ламанця від градового бою ,	бреш	. Переклад розділів XXI-XXXV , записаний у трьох окремих
прилетів до нього посеред зими як інтерв'юер , відряджений від	кисильовського	річника « Україна » ... приведений під його двері , я
спудей з каламарем при поясі й мавританською кривулею на	шабельтасі	, цей чоловік чи то гульбигого (Санчо Панса) , чи то
плоеному комірці : ти задмухнув свічки над своїми летючими , «	чудородними	» , як сновидіння , сторінками . Як мені було добувати
дописувати за тебе твоїм гусячим пером ? Та - чуєш ! -	окульбачено	вже Росинанта , обладовано Сірого , і в сідлі вже їхні
столом французькою мовою книжечку - Мішель Дюїно « Шукачі	фараонії	» і почав її перекладати - угода з видавництвом « Молодь
заступника саме вчасно , щоб стати мовним рефері в цьому	кікбоксі	на Кірова . Зауважу , що сам винуватець на ринзі був відсутній
мої перші , часом дуже ризиковані , па ... Навпроти - це	англізований	з вулиці Енгельса джентльмен , для якого нема літератури
, як сказав би наш великий земляк , такий рід оджежі був	неприсвоєний	, і ноги його роз'їжджалися , і голова наче не держалася
яке справляв на оточення цей роздягнутий і простоволосий	окулярний	. Може , це несвідома людська реакція на появу між нас
березі незамерзаючого моря . Ував , як сніг на голову ,	перебаранчив	працю над Сервантесом , чіпляється , як шевська смола

Використання операторів I, АБО, НЕ

Для одного токена можна задавати декілька альтернативних умов. Усі ці умови можуть розташовуватися всередині єдиної пари квадратних дужок, і між ними ставляться логічні оператори:

& – I

! – НЕ

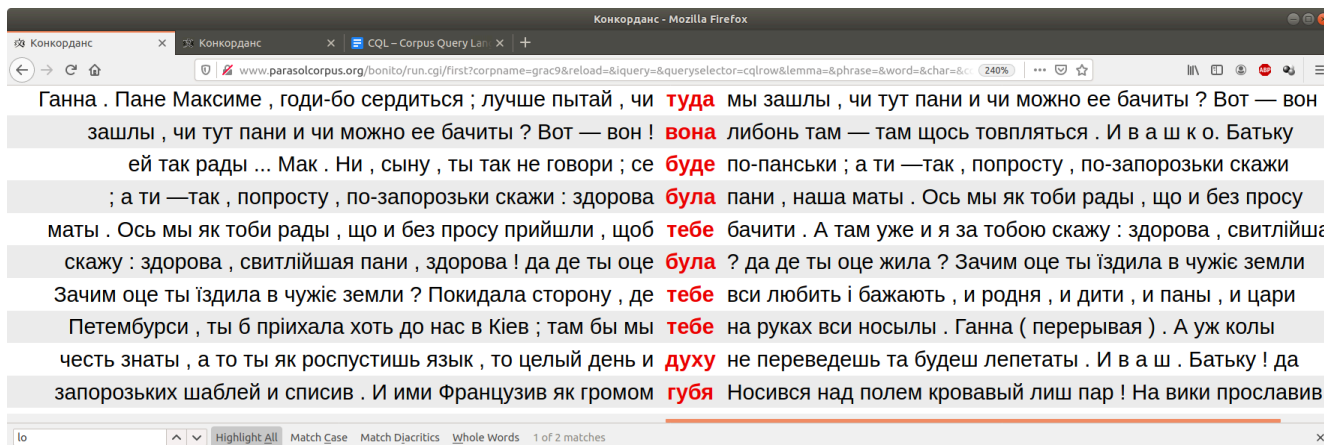
| – АБО

[] – квадратні дужки, що містять набір символів, еквівалентні оператору АБО між цими символами

CQL	Результат
[lemma="мати" & tag="noun.*"]	форми лексеми <i>мати</i> , які мають теґа <i>noun</i>
[lemma="мати" & tag!="verb.*"]	форми лексеми <i>мати</i> , які не мають теґа <i>verb</i>
[lemma= "(y v)вімкнути"]	форми лексем <i>увімкнути</i> і <i>ввімкнути</i>

Запит для пошуку словоформ з 4 літер: голосний – приголосний – голосний – приголосний

[word= "[бвгґджзклмнпрстфхцщ][аоуеієяюі][бвгґджзклмнпрстфхцщ][аоуеієяюі]"



Використання оператора АБО з токенами

Використання АБО між токенами	Використання АБО в межах значення токена	Коментар
[lemma="собака"] [lemma="вовк"]	[lemma="собака вовк"]	При використанні прямого слеша в межах значення токена пошук відбувається швидше
([lemma="великий"][lemma="собака"]) [lemma="вовк"]		При багатослівному пошуку з оператором АБО єдиним можливим способом є зазначення лексем у дужках та використання прямого слеша між ними.

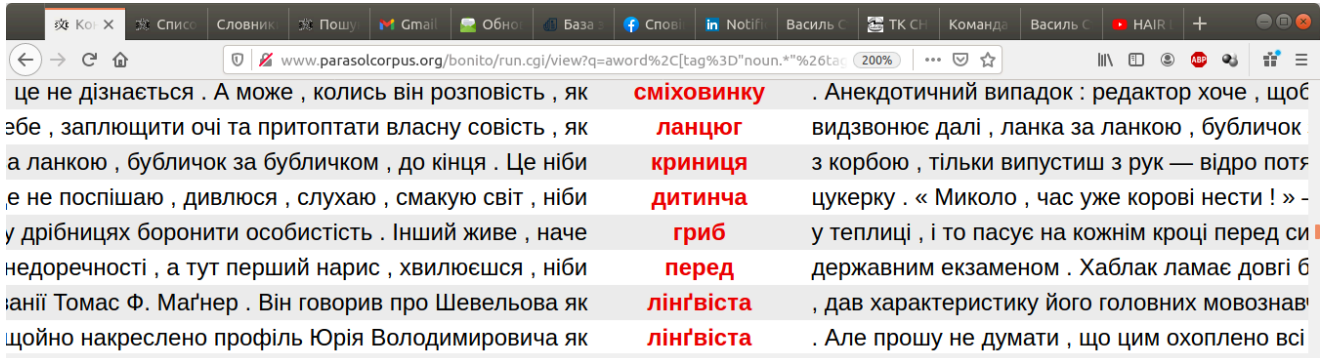
[lemma="собака|вовк"]

Для великих структур у конкордансі буде показано й виділено кольором лише 100 перших токенів.

Виділивши частину запиту як ключові слова за допомогою оператора *within*, можна побудувати частотний список саме за цією частиною запиту.

Наприклад, для пошуку іменників, вжитих після порівняльних сполучників, треба здійснити пошук порівняльних конструкцій (порівняльний сполучник + іменник) і виокремити з них іменники за допомогою оператора *within*:

`[tag="noun.*"] within [lemma="ніби|наче|як|мов|немов|немовби|нібито"] [tag="noun.*"]`

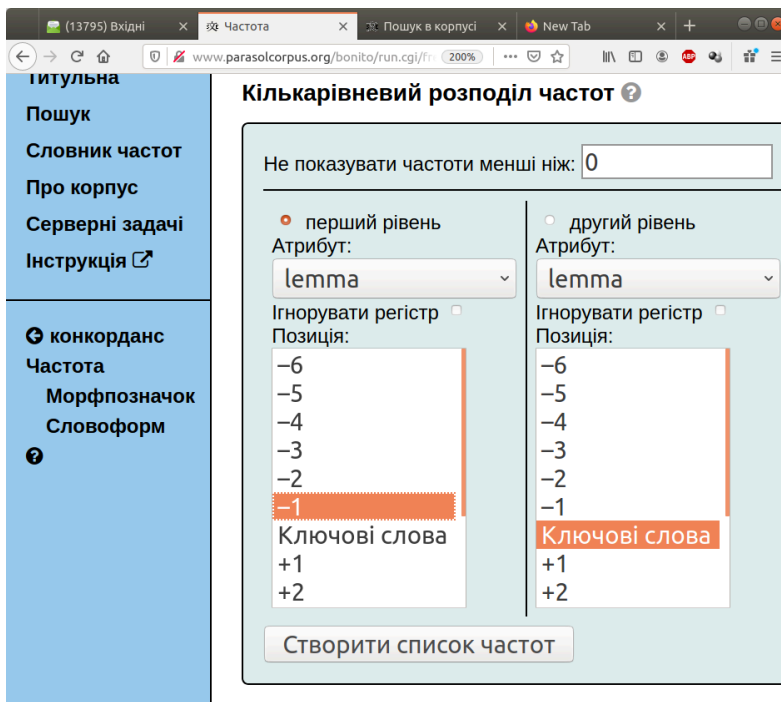


(*Примітка. У ВЕСУМі тег *noun* мають не тільки іменники, а також особові займенники, тег *adj* – прикметники і присвійні займенники. Тому для пошуку іменників без займенників треба уточнити запит: `[tag="noun.*"&tag!=".*pron.*"]`. Для пошуку іменників без займенників: `[tag="adj.*"&tag!=".*pron.*"]`.)

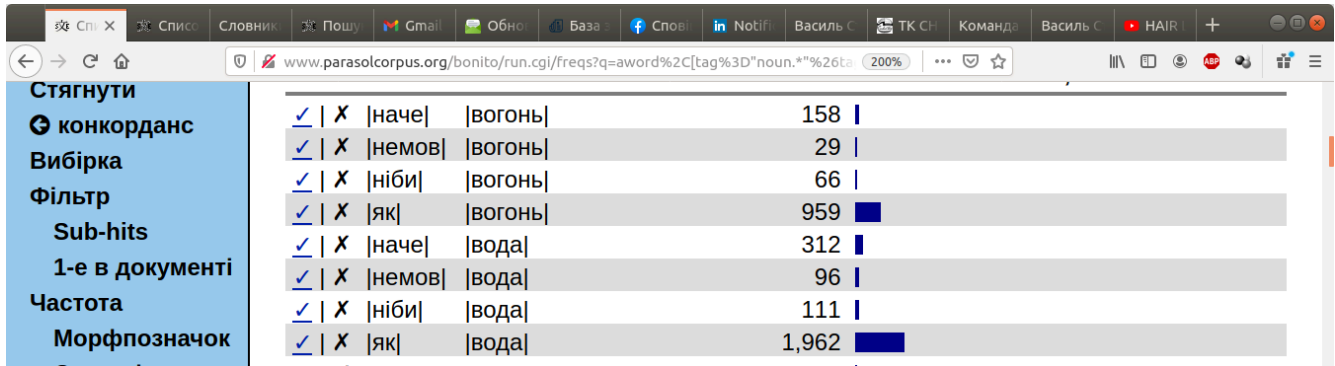
Після цього можна зробити частотний список (меню *Частота*).

Або простий, за ключовим словом, щоб отримати частотний список самих іменників.

Або багаторівневий, щоб побачити сполучення слів. Для цього в меню *Частота* слід обрати на першому рівні лемму попереднього токена по відношенню до ключового слова (позиція -1), на другому рівні – саме ключове слово і створити список.



Результат:

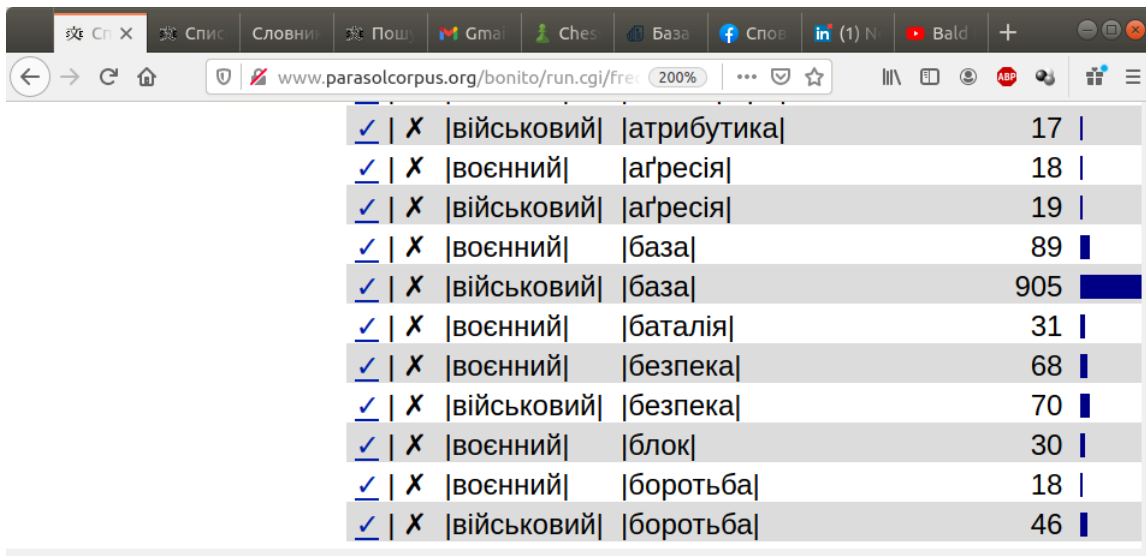


✓ X	наче	вогонь	158
✓ X	немов	вогонь	29
✓ X	ніби	вогонь	66
✓ X	як	вогонь	959
✓ X	наче	вода	312
✓ X	немов	вода	96
✓ X	ніби	вода	111
✓ X	як	вода	1,962

Аналогічно можна дослідити частоту поєднання лем *військовий* і *воєнний* з різними іменниками.

```
[tag="noun.*"] within [lemma="військовий|воєнний"][tag="noun.*"]
```

Дворівневий список частот можна сортувати за першим компонентом, другим компонентом або частотою. Тут список посортовано за другим словом:



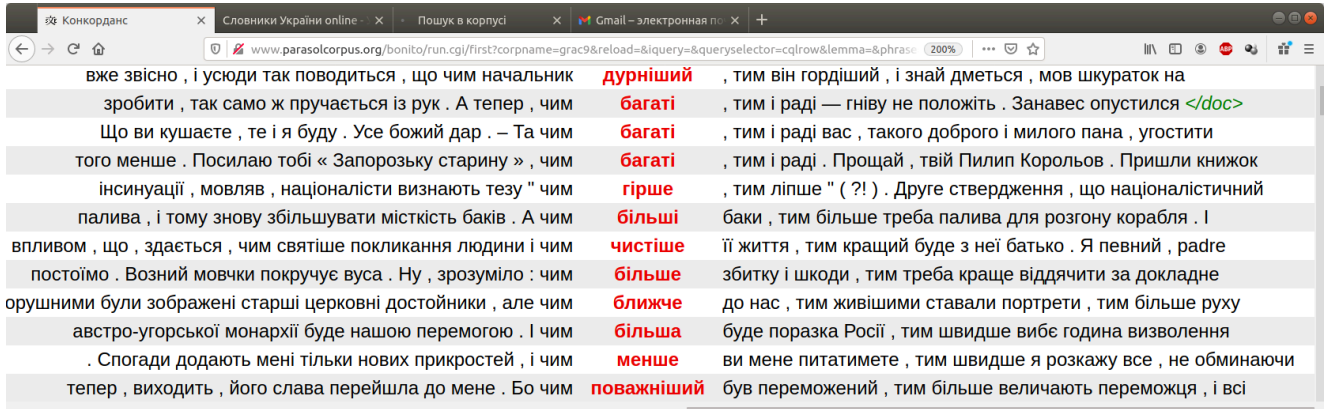
✓ X	військовий	атрибутика	17
✓ X	воєнний	аґресія	18
✓ X	військовий	аґресія	19
✓ X	воєнний	база	89
✓ X	військовий	база	905
✓ X	воєнний	баталія	31
✓ X	воєнний	безпека	68
✓ X	військовий	безпека	70
✓ X	воєнний	блок	30
✓ X	воєнний	боротьба	18
✓ X	військовий	боротьба	46

Вкладення («гніздування») при використанні структур

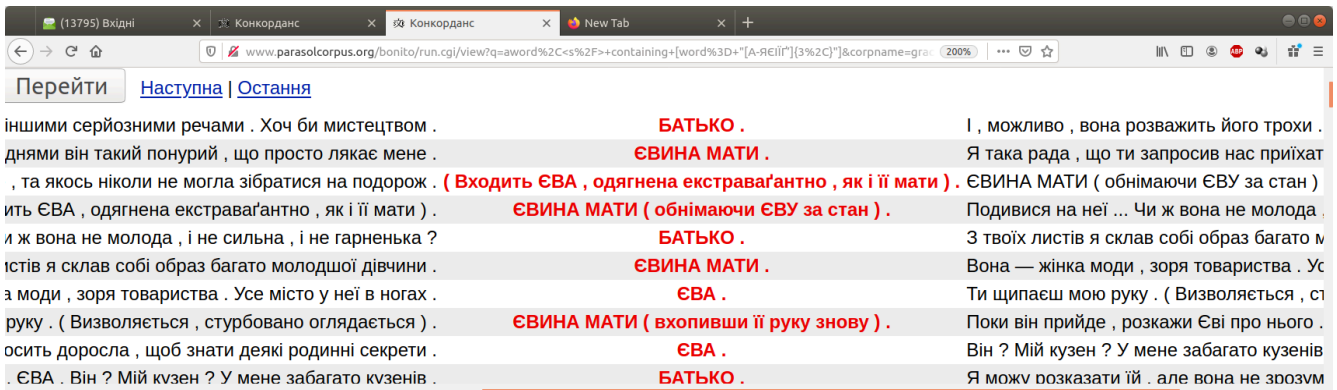
За необхідності оператор *within* можна вживати кілька разів в одному запиті.

Наприклад, запит для пошуку прикметників (без присвійних займенників) у межах конструкції, що починається зі словоформи *чим* і закінчується словоформою *тим*, між якими може бути ще 0-5 токенів, і все це знаходиться у межах одного речення:

[tag="adj.*".*pron.*"] within [word="чим"]{0,5}[word="тим"] within <s/>

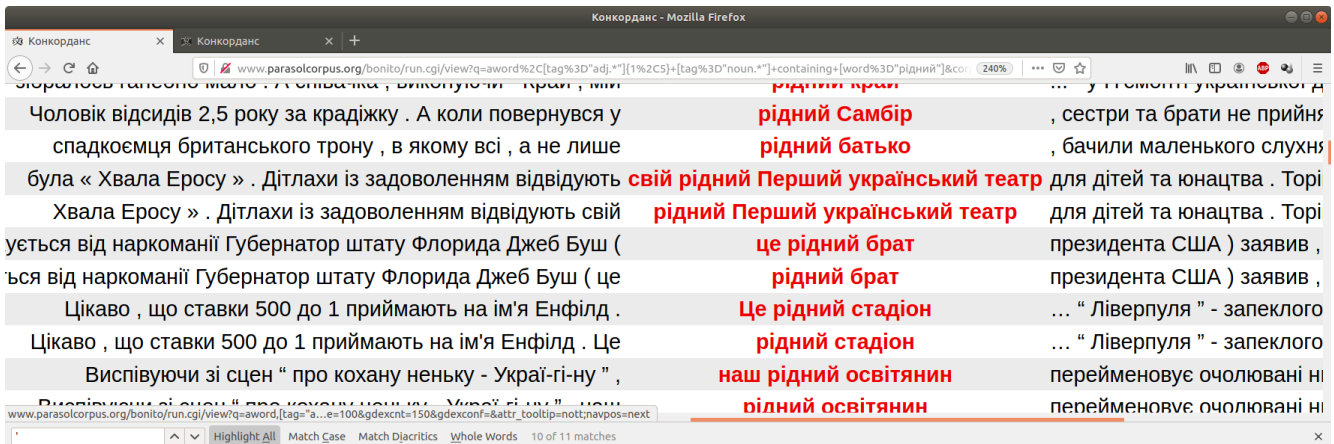


Цей вираз знайде всі речення, у яких зустрічається слово з 3 і більше великих літер:
<s/> containing [word= "[А-ЯЄІЇҐ"]{3,}"]



За допомогою цього запиту здійснюється пошук іменникових словосполучень (послідовності із максимум 5 прикметників та іменником), які містять прикметник "рідний".

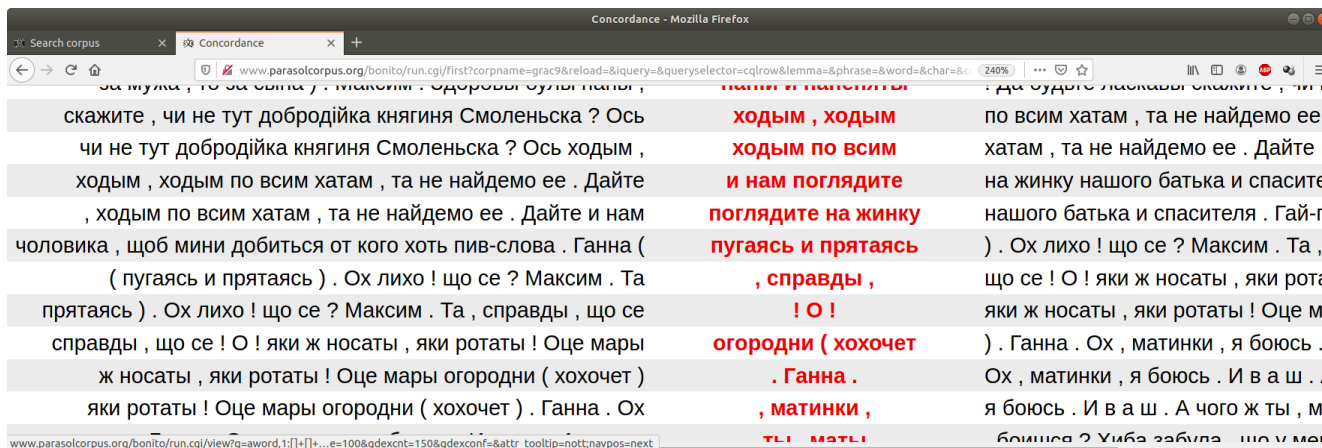
[tag="adj.*"]{1,5} [tag="noun.*"] containing [word="рідний"]



Цей вираз має здійснювати вибірку усіх повторених лем на відстані через один токен:

1:[] [] 2:[] & 1.lemma = 2.lemma

Однак, на практиці лемми іншомовних вкраплень та пунктуаційних тегів позначені однаково, тому результат буде мало корисним:

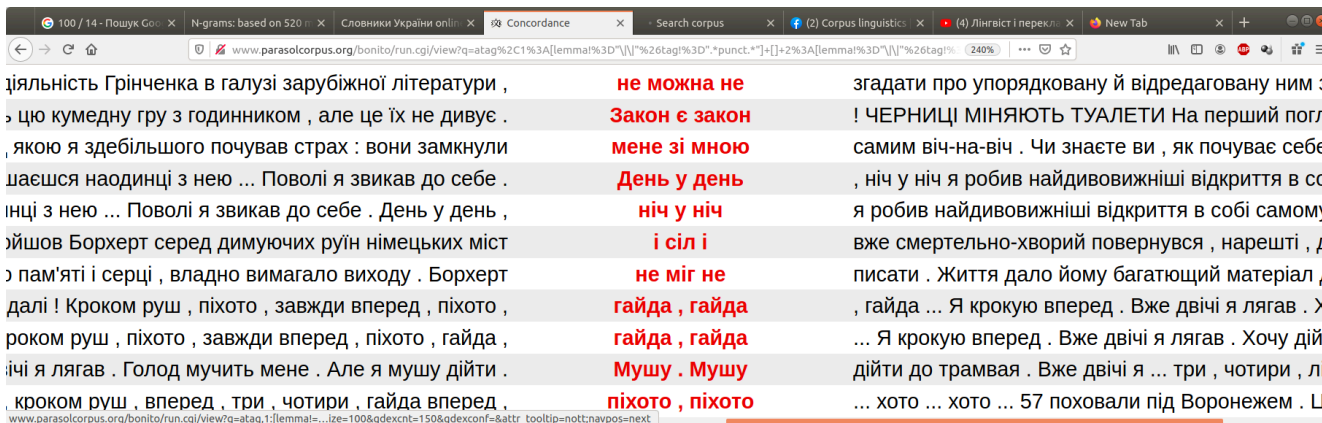


Щоб позбутися нелематизованих, чужомовних лексем та знаків пунктуації у вибраних виразах, необхідно додати всередині токена умови.

Наприклад, що лема складається з українських літер:

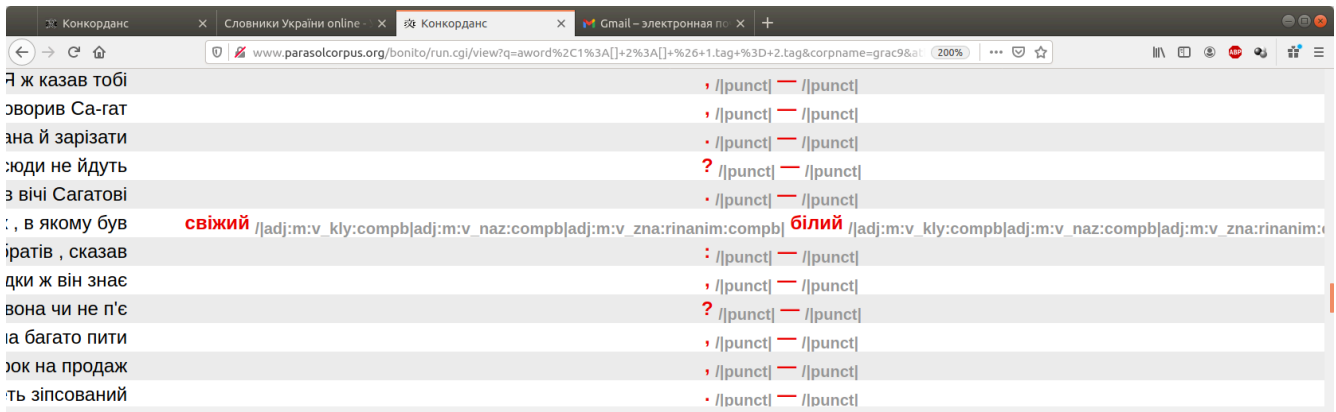
1:[lemma="[а-яієґ'-].*"] [] 2:[] & 1.lemma = 2.lemma

Додавання подібних умов після амперсанда не спрацює.



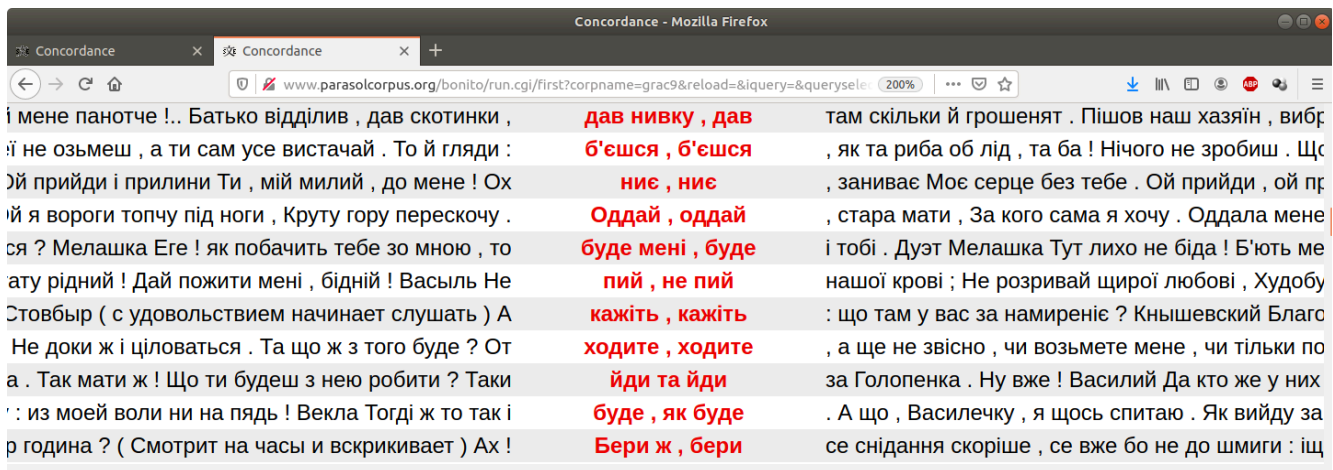
Цей запит спрямований на вибірку будь-яких двох словоформ, що стоять поруч і мають однаковий набір тегів, тобто два іменники, два прикметники, два дієслова в однакових граматичних формах (часто це будуть однорідні члени речення або ж два підряд знаки пунктуації):

1:[] 2:[] & 1.tag = 2.tag



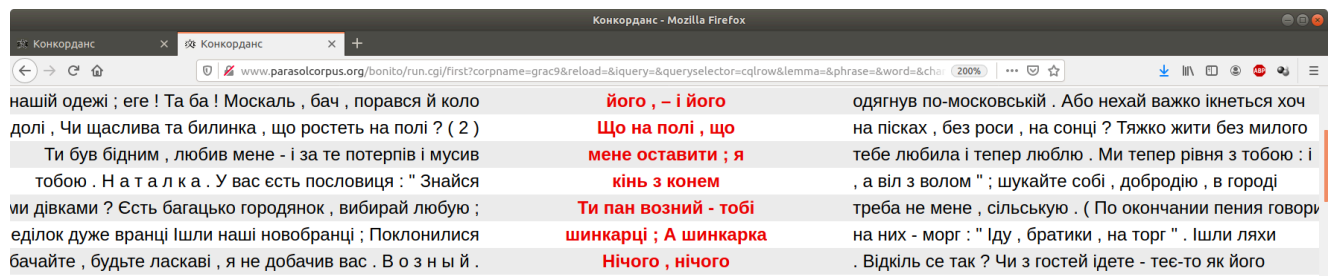
Наступний запит знаходить вирази з повтореними дієслівними лексемами в межах одного речення:

1:[tag="verb.*"]{0,2}2:[tag="verb.*"] & 1.lemma=2.lemma within <s/>



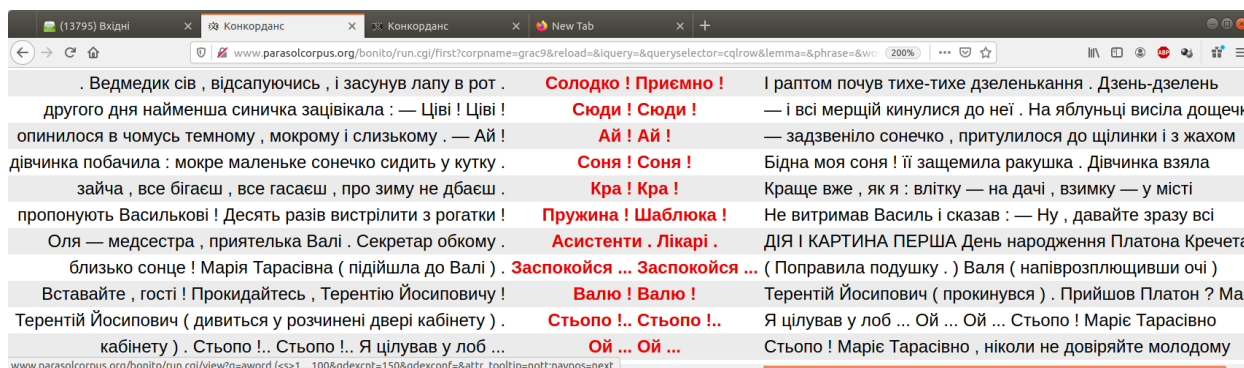
Використання операторів *containing* чи *within* з глобальними умовами може вимагати круглих дужок навколо першої частини запиту. Необхідність та перевага дужок полягає в чіткому визначенні пріоритетності операторів: спочатку виконують операції, що в дужках.

(1:[tag="noun.*"] []{0,3} 2:[] within < s/>) & 1.lemma=2.lemma



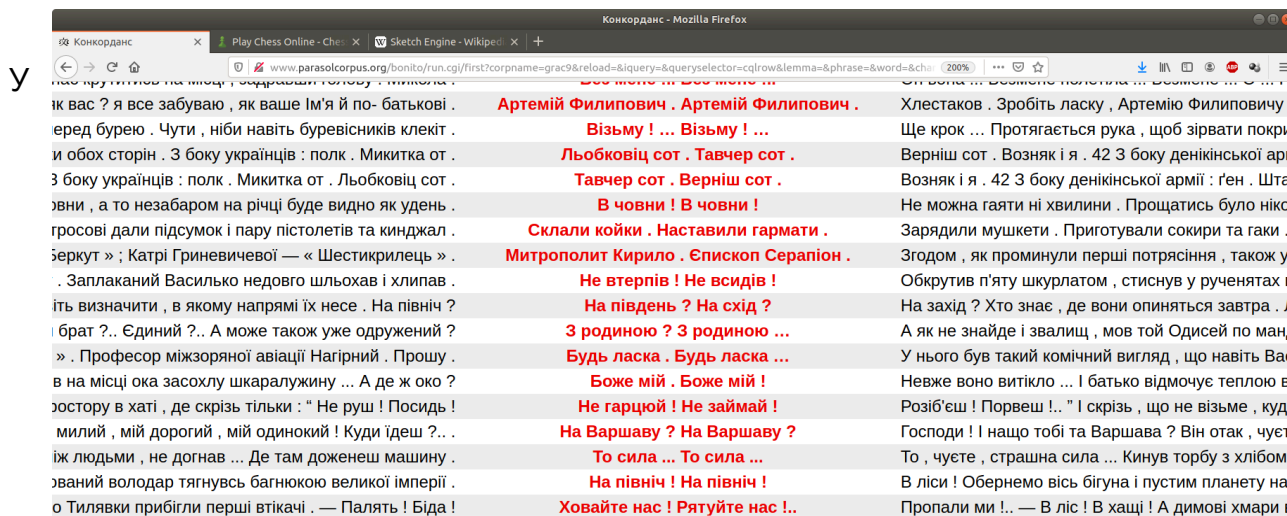
Наступний запит дозволяє знайти мінімальні (однослівні) паралелізми (речення з подібною синтаксичною структурою), якщо припустити, що речення з рівнотегованими токенами мають подібну синтаксичну структуру:

(<s>1: [] 2: [] </s>)(<s> 3:[] 4:[] </s>) & 1.tag=3.tag & 2.tag=4.tag



Аналогічно, двослівні паралелізми можна знайти за таким запитом:

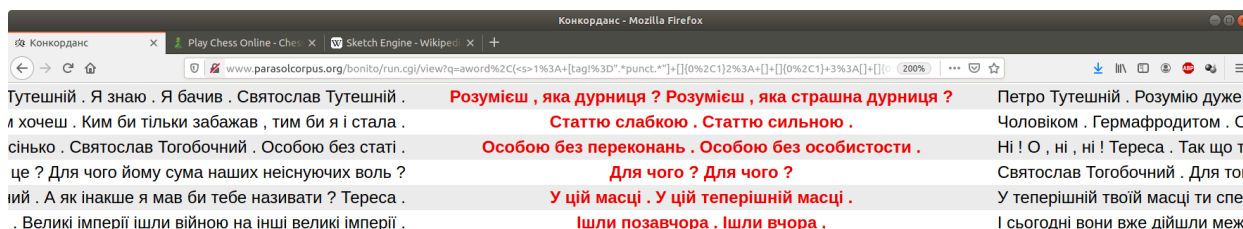
(<s>1: [tag!="punct"] 2: [] 3:[]</s>)(<s> 4:[] 5:[] 6:[] </s>) & 1.tag=4.tag & 2.tag=5.tag & 3.tag=6.tag



цьому запиті

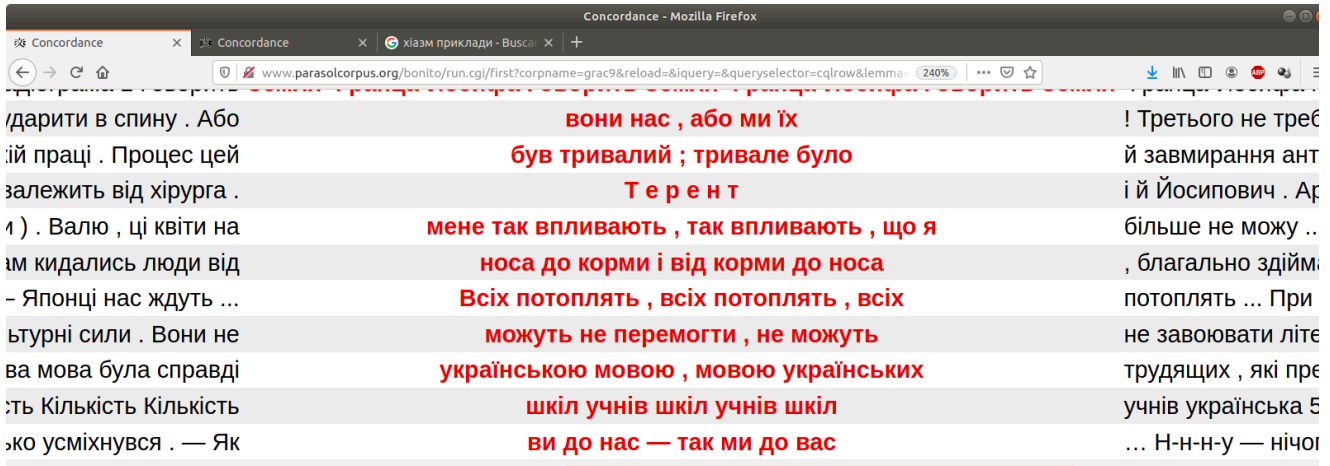
(<s>1: [tag!="punct"] []{0,1}2: [] []{0,1} 3:[] []{0,1}</s>)(<s> 4:[] []{0,1} 5:[] []{0,1} 6:[] []{0,1} </s>) & 1.tag=4.tag & 2.tag=5.tag & 3.tag=6.tag

також враховано, що паралелізми не обов'язково передбачають ідентичні структури речень, а лише подібні, тому й однаковими тегами мають характеризуватися не всі токени, а переважна частина з них. Між токенами, що явно збігаються в обох речень, можуть розташовуватися й інші токени, що відрізняються:



Наступний запит дозволяє знайти деякі хіязми (стилістичної фігури, заснованої на хрестоподібному застосуванні слів А-В-В-А). У хіязмі мають виконуватися три формальні умови: країні лексеми однакові, серединні лексеми однакові, крайні і серединні лексеми неоднакові:

```
1:[lemma="[a-яіієr"-].*" & tag!="conj.*|prep"] []{0,2} 2:[lemma="[a-яіієr"-].*" & tag!="conj.*|prep"]
[]{0,2} 3:[] []{0,2} 4:[] & 1.lemma=4.lemma & 2.lemma=3.lemma & 1.lemma!=2.lemma within
</s/>
```



Частотність з глобальними умовами

Для визначення частоти певної структури використовується функція f(), а сама структура заноситься в дужки.

Запит для пошуку лем з частотою 1:

```
1: [] & f(1.lemma)=1
```

Те саме без складних слів, номерів і іншомовних слів:

Результат пошуку за таким запитом у власних текстах і перекладах Анатолія Перепаді:

```
1: [tag!="number"&lemma!=".*-.*"&tag!=".*foreign.*"] & f(1.lemma)=1
```

– заперечив старшина ,	смілішаючи	, коли небезпека мину
идило мені більше , ніж	ослизлість	медуз , які каляли мен
розмаїття меблям , ніби	відлущуючи	квітучий шовк фотелів
екинулася наймилішим ,	найгречнішим	хлопцем , якого я тілы
, здавався їй уже якимсь	лжемаркізом	. Вона виявляла таке с
ка ; моє минуле , одразу	потьмарившись	, уже не відкидало пер
і спершу взяв за коханок	велоперегонників	або чемпіонів з боксу ,
п поставала темноволоса	гольфістка	, хоч як би вони різни
еже своє життя , чинять	якнаймудріше	, бо хоча я особисто не
та така сама нелегка , як	виїжджування	коня , і така сама засп
що вона майже весь час	#була	не в гуморі , і це її нев
дно , каравела , галеас ,	буцентавр	[189] » . Альбертина :
шою гостротою відчуває	просяклість	пахощами квітів та плс
леліє світляна блакить ,	заслонювана	віддзеркаленим хвиля.
, буде людський голос у	фототелефоні	потомних літ : у його з
ків . Так ось , ця дамуля	акторувала	, але в акторстві їй не
ий апофеоз зворушено ,	протверезуючи	себе думкою , що Безс
замість нашого ока буде	найматеріальніша	лінза , фотографічна п
ріус , який був секретар	антиревізійністського	комітету . – Тепер каж
уть , задля певних слідів	неогомерівської	(вже занедбані) ман
рові знаєте , страшенна	рутинерка	, вона сахається всьог
к пишеться і в книзі) : «	прочисні	» . Котгар , на превели
ве з бабусі знято п'явки ,	конгестія	вернулася з новою сил
юру , а навіть ще більше	гоноруючись	: бачте , який я , відом
надцять ! Альбертинина	вилюднілість	виявилася ще виразні
ршим мені впадав у вічі	компрометовний	лист , так само як він і
трів , навіть улітку часто	обгортуваний	млою , саме нині , у не
дну з вулиць названо від	чорноризок	капуцинок , чий кляшт
ли : « Боже свідче , ну й	кумедниця	наша Оріана ! Вдає ког

Якщо здійснити вибірку сполучень з двох іменників в однакових граматичних формах, перший з яких вживаний значно частіше, ніж другий, можна видобути багато прикладів обмежувальних прикладок, оскільки перша лексема в словосполученні буде переважно ширша за значенням, ніж друга:

```
1:[tag="noun.*"&tag!=".*pron.*"] 2:[tag="noun.*"&tag!=".*pron.*"] & f(1.lemma)>5000 & f(2.lemma)<100 & 1.tag=2.tag
```

Григорій К...	Безверхому хуторі , біля панських хат , дивлячись , що	панна хорунжівна	, Олена Йосиповна , сидить на посаді з паном су
Григорій К...	, на яких би то м'якесеньких подушечках спав би він з	панною хорунжівною	, і усе ... і тее ... а тут довельсь лягати на своїй пи
Григорій К...	що й полетів еси , аки птиця пернатая ; вона обуяла і	панну хорунжівну	, нині паню Халявську ; та ще - ох ! - і сотничку , і
Григорій К...	змінили . А його жінка , себто Олена Йосиповна , що була	панна хорунжівна	і попереду жила у Безверхому хуторі , що на Сух
,3 Лобізви...	людьми мов з звірами . А що ще нас більше дивує , що і	панна експедиторка	, котра славиться в околиці взірцем всіх жіночи
Михайло Шо...	під ноги , спашують назря або сплять . Хрѡпнувши об	коліно граблище	, він поскреготав зубами , поніс рептух із сіном , і
Оксана Іва...	здавалося , що то справжній ліс . Там жили підстаркуватий	дядько дятлик	і стара бабуся сова — серйозні , помірковані пта
Оксана Іва...	вона вилітала на лови , а вдень спочивала . А працюватий	дядько дятлик	, як завжди , був зайнятий . Він стукав по стовбу
Оксана Іва...	і лиха бути не може . — А , так там синички ! — сказав	дядько дятлик	. — То хай вони ' й доглядають яблуньку ! От зак
Оксана Іва...	— Тікайте ! Тікайте швидше ! — почули вони . То кричав	дядько дятлик	, у якого , нарешті , була перерва в роботі . Вмит
Оксана Іва...	що вона могла зробити ? Вона тільки ледь-ледь зітхала .	Дядько дятлик	спокійно собі працював у старому саду і не хотів
Оксана Іва...	дітей , які не забувають щодня виносити їм їсти . Старий	дядько дятлик	недовіжливо похитав головою , але пересвідчивс
Оксана Іва...	добре знали свою любиму яблуньку . Незабаром прилетів	дядько дятлик	, зосереджено постукав дзьобом по стовбуру . —
Оксана Іва...	Злякані і схвильовані вкрай синички швидше подалися до	дядька дятлика	. — Яке нахабство ! — обурився він . — Негайно
Олексій Но...	Ітуруп Курильського пасма . Він розпорядився лягти на	курс норд-вест	17° . Туман густішав . Людям з містка нічого не б
Олексій Но...	англійського пароплава " Айко-на " . Зараз же " Терек " ліг На	курс зюйд-ост	45° , маючи попереду " Айко-ну " , щоб відійти в і
Олексій Но...	заступив його , тяжко поранений . О третій годині лягли на	курс зюйд-вест	48° і пішли восьмивузловим ходом , направляюч
Олексій Но...	вона була виконувати роль озброєної яхти для великого	князя генерал-адмірала	Олексія Олександровича . Будівники її мало зве
Сергій Пил...	Шовіністичний чад не захопив його . Навпаки , він з групою	робітників соцдемократів	складає статті проти війни , і Самобитнік випуск
Володимир ...	по-особливому радіоактивне , якщо інтенсивно реагує на зміну	частот радіоколивань	. А воно ж саме так , коли правильно те , що ви р

Пошук лівого і правого контексту – оператор *meet*

Пошук лівого та правого контексту здійснюється за допомогою оператора *meet*.

Цей запит скерований на пошук випадків вживання лема *музикальний*, якщо на відстані до двох токенів зліва або до двох токенів справа від неї є лема *голос*:

```
(meet[lemma="музикальний"][lemma="голос"] -2 2)
```

а спів . То співала на голоси **музикальна** молодь містечка . На рунду
ний , мелодійний » : « Голос **музикальний** » (М. Коцюбинський) . От
й музичний вечір » , « голос **музикальний** » . Недостойний і негідний
а , – він робився чарівним ; **музикальний** голос його переливав у сер
> , « приємний звучанням » (**музикальний** голос) . « Він гарний ... його
здаються свіжішими ... Голос **музикальний** ... » (М. Коцюбинський .) М
дбігла до письмового стола і **музикальним** ніжним голосом , трохи гун
даються свіжішими ... Голос **музикальний** ... Ти його побачиш , ти му
е , лагідна , ніжна , тиха , її **музикальний** голос помагав мені думати
е , лагідна , ніжна , тиха . Її **музикальний** голос помагав мені думати
рім звичних звуків голосу та **музикального** обрамлення , ще є й лаконі
ї , що мають добрий голос і **музикальний** слух . Першенство мають в

Лише перша частина, тобто [lemma="музикальний"], виділяється як KWIC в конкордансі, друга частина, [lemma="голос"], використовується як фільтр.

Поєднання двох запитів з оператором *meet* – оператор *union*

Оператор *union* призначений для поєднання результатів двох запитів оператора *meet* і відображення їх в одному конкордансі. Оператори *meet* та *union* мають стояти всередині круглих дужок.

```
(union (meet [lemma="над"] [lemma="під"] -2 2) (meet [lemma="понад"] [lemma="попід"] -2 2))
```

Цей CQL запит знайде:

- усі лема *над* з лемою *під* на відстані до 3 токенів зліва чи справа;
- усі лема *понад* з лемою *попід* на відстані до 3 токенів зліва чи справа;
- поєднає результати першого запиту та результати другого запиту та відображатиме їх в одному конкордансі, лема *над* і *понад* будуть виділені в центрі як ключові слова.

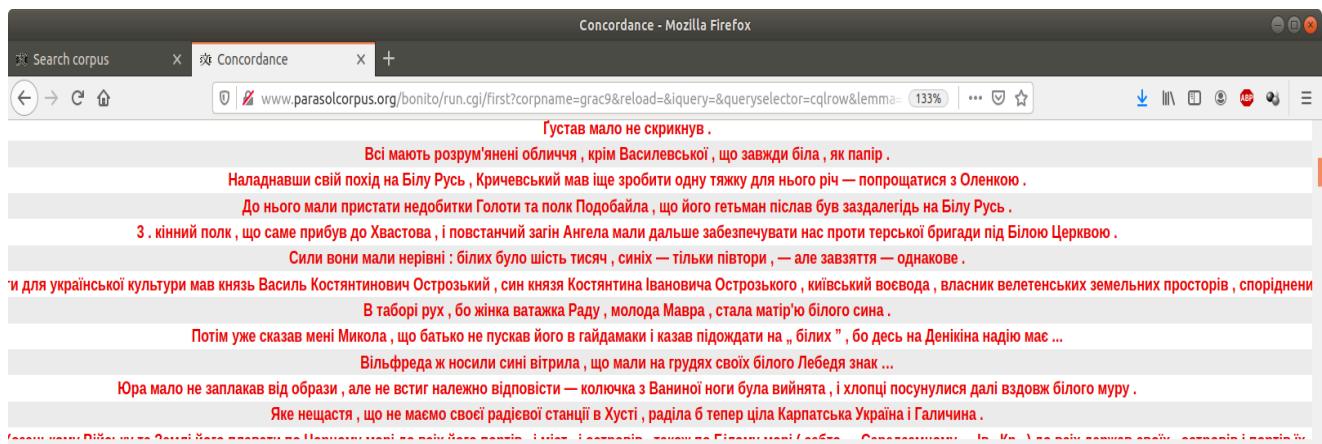
, і я побіцяв . Там по середині баржі під стелею	над	верхніми нарами є щілина завширшки більше десяти
частина лісом пішла на Гуту , а частина попід лісом	понад	Сівчину на Бірчу . Два озброєні поляки підбігли до
лом у Франції звістку про перемогу царя під Полтавою	над	Карлом XII і гетьманом І. Мазепою зустріли з недовірою
" Дума українна " Адама Чагровського , " Перемога	над	татарами під Теревовлею 1524 року " М. Гусовського
зцем року 1512-го " , поема М. Гусовського " Перемога	над	турками під Теревовлею 1524 року " , дві промови-тракта
ішли до висновку , що краще сидіти в тюрмі під дахом	над	головою і мати постійно хоч який кусок хліба , аніж
Львова , чотовий , згинув в бою 3 червня 1947 року	над	Команчею під час рейду на захід . Веремчук Аркадій
, авто , парканчик , туди , де вона , під навислі	над	тротуаром будинки , магазини та липи-дерева . Вони

Зауважте, що кожен з операторів *meet* та *union* повинен бути всередині своєї пари дужок.

Оператор *union* працює тільки в поєднанні з *meet*. Для поєднання результатів інших запитів, що не містять *meet*, можна скористатися логічним оператором АБО (вертикальний слеш |).

Оператори *meet* та *union* можуть використовуватися, коли потрібно знайти позиційний атрибут (тобто декілька квадратних дужок) і можуть поєднуватися з операторами *within/containing*. Наприклад:

```
< s/> containing (meet [lemma="мати"] [tag="*.prop.*"] -3 5) containing (meet [tag="noun.*"] [lemma="білий"] -4 2)
```



Структури

Речення й документи в ГРАКу марковані як структури (<s> та <doc> відповідно), що дозволяє врахувати їх у запиті.

Кожна структура може мати значення, наприклад, документ позначений датою, коли він був написаний, прізвищем автора тощо; ці значення також можна використати як критерії пошуку.

Початок структури

Для позначення початку структури слід використати: <s> або <doc>.

Запит для пошуку лема *природа* на початку речення:

<s> [lemma="природа"]

Корпус: Gras v.8

Простий запит: Побудувати конкорданс

[Типи запитів](#) [Контекст](#) [Типи текстів](#)

Тип запиту: простий лема фраза слово символ CQL

Лема:

Фраза:

Словоформа: добрати реєстр

Символ:

CQL: Автоматичний атрибут: lemma

[Система позначок CQL-конструктор](#)

Побудувати конкорданс Очистити

звідчування . **Природа** та її життя прекрасні , але людина стала окремо , ніж Олесь ... **Природа** в Олесь дуже абстрагована – ідеалізована , в раниців , любові . **Природа** ще не розгорнула всіх див своїх ; вона тільки що ватись самі . **Природа** дає для людської національної творчості тільки сієми в другу . **Природа** так само не продукує грошей , як і банкірів чи векто оточують . **Природа** - ж електричної сили по своїй суті відрізняється а » природа ? **Природа** скрізь інтернаціональна . Це самостійницька книжка щаслива . **Природа** цвіла і сяла , а неосяжне , завжди спокійне море / бика , шиї . **Природа** не окупувала , по-первісному щедро відливаючи цманне море . **Природа** , цей незрівнянний художник , шукала цієї ночі хильної корови . **Природа** обмежилася сокирою , створюючи його , а щоб змилій ерос ... **Природа** не терпить винищення – вона намагається негайно циро вірили . **Природа** ! Ліс , безмежний густий ліс , де непорушно стоять цієї кишені . **Природа** могла б , за цим кравецьким здогадом , заощадити чному горлі . **Природа** не обдарувала Степана співучою здібністю його на здійснення . **Природа** замкнула свої вигідні притулки , але жоден дощ не вго злитися . **Природа** для нас – спогад і спочинок . – Це ліси й поля ? Нашим тощо . **Природу** цього володарського " ми " дуже влучно схарактерна природу . **Природа** природою , але ж жінка не весь час зв'язана приресподарства . **Природа** мало турбується про нас і ніколи не розподіляє дооги , лапи) . **Природа** , ніби-то з економії , не знищила зовсім зябер , . стилів . 3 . **Природа** в музиці . 4 . Про творчість окремих композиторів виходящих ! **Природа** наче б то існує вже й давно , А діло . . жах ! – жінці – час . **Природа** одібрала право це у скопців та гермофродитів . Одсі без кінця . **Природа** чекала великого свята і врочисто вбиралась . Мої віт вимирає . **Природа** обернулася на продукт фабричного виробництва .

Запит для пошуку леми *бути* на початку твору:

<doc>[lemma="бути"]

Твій Іван </doc><doc> **Був** май 1880 року ; були перші дні мая , але над Кишиново
ивабите ! </doc><doc> **Було** колись . За давніх-давен було . За Тараса Бульби
з Заходу . </doc><doc> **Є** анекдот про двох мужиків , що побилися на базарі
злочину . </doc><doc> **Будьмо** вірні традиції ! З приводу Пастирського Листа їх
і Долині ! </doc><doc> **Була** провідна неділя і розсміяне пополудне , і верби маїлис
іл і міст . </doc><doc> **Був** початок червня 1940 р. На світі Божому правдоподібно
ю 13 га . </doc><doc> **Був** учнем , став учителем Приїхав я в Донбас по зову
ицюком . </doc><doc> **Будемо** і надалі передовими Днями , в клубі ім. Ворошилова
зварису ! </doc><doc> **Бути** донором - почесна справа Дорогоцінна кров радянськи
мічників . </doc><doc> **Бути** активним громадсько-політичним діячем Радянської
волення . </doc><doc> **Будемо** боротись за високий врожай цукрових буряків Один
айомила . </doc><doc> **Є** приємна втома Є приємна втома , неспокійна , Коли
югунець . </doc><doc> **Є** старовинна приказка : « Не все те золото , що блищить
днішими . </doc><doc> **Був** собі один учитель , який жив у Капелюсі Винограду
області . </doc><doc> **Є** НА СВІТІ ОДНА « ПРАВДА » (З народної мудрості)
елений ... </doc><doc> **Будьмо** разом Будьмо разом , як ті рослини , Що сплелися
я богам . </doc><doc> **Було** нас разом 25 студентів з Франції . Ми їхали через
серця » . </doc><doc> **БУВ** ТРАВЕНЬ Неповторний , найщасливіший і найблагоденств
шканців . </doc><doc> **ЄСТЬ** , СТОЯТИ НА СМЕРТЬ ! ДОРОГА Сиза зловісна хмара на
і корови . </doc><doc> **Є** тепла конура і добра страва , А він гарчить . У чому
дуть цілі . </doc><doc> **Є** в нього нюх , і бачить він далеко — Скоріш за всіх
ячих зір . </doc><doc> **БУЛА** була . була і відчувала . була і відчувала вістря
а станції . </doc><doc> **Є** СТІЛЬКИ КРАЇВ Є стільки країн , є стільки країв ,
рекладу . </doc><doc> **Була** собі в гаю хатка , а в тій хатці Жила Жінка з сином
в її лоні . </doc><doc> **Будь** , як земля , тверда і вірна . Як статуя , заціпеніла
лування . </doc><doc> **Є** письменники , власні долі яких не дуже відбиваються

Закінчення структури

Для позначення закінчення структури слід використати:

</s> або </doc>

Закінчення речення: </s>

Корпус:

Простий запит:

[Типи запитів](#) [Контекст](#) [Типи текстів](#) ⓘ

Тип запиту простий лема фраза слово символ CQL

Лема:

Фраза:

Словоформа: добирати реєстр

Символ:

CQL: Автоматичний атрибут:

[Система позначок](#) [CQL-конструктор](#)

Запит 32,508,045 (64,261.00 на мільйон) | взято перших 10,000,000 рядків (взяти вкладкові) ⓘ

Сторінка з 100,000

Кравов'як	великим святом , земляки з перемогою , товариші земляки !	Дуже вдячний за численні вісті , які одержав від вас у
Кравов'як	Шумска і що допомагали в цих злочинах німцям бандерівцям .	Звірлячі справи творили ці насильники на вашій землі над
Кравов'як	ці насильники на вашій землі над нашими мирними людьми .	Гнів роздирав моє серце і робив нещадним у завершених
Кравов'як	роздирав моє серце і робив нещадним у завершених болях .	Тільки радість перемоги вгамовує горе . В рядках тих ,
Кравов'як	завершених болях . Тільки радість перемоги вгамовує горе .	В рядках тих , хто відстоює нашу Батьківщину , приніс свободу
Кравов'як	незалежність пригніченим народам , немало наших земляків .	Вони - учасники здійснення великої визвольної місії Червоною
Кравов'як	учасники здійснення великої визвольної місії Червоною Армією .	Скоро вони повернуться в свої сім'ї . Не сумніваюсь ,
Кравов'як	Червоною Армією . Скоро вони повернуться в свої сім'ї .	Не сумніваюсь , що ви приймете , як сина , в свою велику
Кравов'як	що ви приймете , як сина , в свою велику сім'ю і мене .	У вас священна програма - швидше відбудувати господарство
Кравов'як	відбудувати господарство й культуру , зруйновані німцями .	Прикладайте ж усіх зусиль для успішного здійснення цього
Кравов'як	ж усіх зусиль для успішного здійснення цього завдання .	Ми , воїни-терноплічани , відплатили гітлерівцям за кров
Кравов'як	слези , за всю кризду й знущання , які вони заподіяли вам .	Один у нас незведений рахунок . Може ходять ще по моїй
Кравов'як	які вони заподіяли вам . Один у нас незведений рахунок .	Може ходять ще по моїй землі бандерівські дітобвищ ,
Кравов'як	бандерівські дітобвищ , що знищили сім'ї багатьох з нас .	Відплатіть їм за наше горе . До скорой зустрічі , дорогі
Кравов'як	знищили сім'ї багатьох з нас . Відплатіть їм за наше горе .	До скорой зустрічі , дорогі земляки ! Запорожець Максим
Кравов'як	їм за наше горе . До скорой зустрічі , дорогі земляки !	Запорожець Максим Голоченко Ганна , жєна его Івашко ,
Микола Гнідич, [П'єса без назви], 1816		

Атрибути структури

Щоб знайти всі документи в корпусі, створені у 2011 році, необхідно виконати такий запит:

```
<doc date="2011">
```

A. Боровик, Чи не настав той час?, 2011

A. В. Береза, Чинники підвищення демократичності управлінських мереж, 2011

A. В. Береза, Громадські ради в Україні: історія становлення, 2011

A. В. Береза, Парадокс посередника у запровадженні електронної демократії, 2011

A. В. Береза, Підходи міжнародних організацій до оцінки ефективності публічного управління, 2011

A. В. Бондаренко, Правосвідомість як предмет соціально-філософського аналізу, 2011

A. В. Вітер & В. А. Доброскок & С. М. Маліченко & та ін., Вплив кліностатування на симбіоз люцерни посівної з ризобіями, 2011

A. В. Євтехова & Е. О. Беспояско & А. В. Злобіна, Мінералогічна зональність покладів гематитових кварцитів Ігулецького родовища, 2011

A. В. Зубенко, Особливості поняття, структури, змісту та форми інтерпретаційних актів, 2011

A. В. Козоріз, Соціалізація ринкової економіки та її значення для державних фінансів України в умовах глобалізаційних процесів, 2011

A. В. Корнилюк, Аналіз розвитку теорій структури капіталу підприємств, 2011

A. В. Коструба, Нормотворчі засоби встановлення правоприпиняючих юридичних фактів у цивільному праві, 2011

A. В. Кофанов & О. Л. Кобилянський & Я. В. Кузьмічов & та ін., Криміналістика, 2011

A. В. Кудимов, По слідам загиблї флотилії, 2011

A. В. Литвиненко & М. С. В'ялик, Актуальні проблеми безробіття молоді в сучасній економіці України, 2011

A. В. Мохонько, Ефективність дії правових та політичних норм, 2011

A. В. Мохонько, Право як соціально-нормативний регулятор, 2011

A. В. Найчук, Генезис громадських організацій: історична ретроспектива, 2011

A. В. Огієнко, Аналіз розвитку зовнішньоекономічної діяльності України в умовах глобалізації світової економіки, 2011

A. В. Огієнко, Аналіз стану та напрямки удосконалення зовнішньої торгівлі України, 2011

A. В. Плига, Літня орнітофауна проєктованого національного природного парку «Подесіння» в заплаві р. Десна, 2011

A. В. Приліпко, Правий статус іноземних підрозділів військово-морських сил під час конфліктів неміжнародного характеру, 2011

A. В. Рогозовська & І. М. Романюк, Історіографія історії середньовічного Поділля в контексті битви на Синіх Водах 1362 р., 2011

A. В. Себко, Управління інформаційними ресурсами підприємств АПК, 2011

A. В. Сотников, Методика проведення експериментально-фонетичного дослідження британських політичних промов, 2011

A. В. Черкасов, Формування системи соціальних стратегій регіону, 2011

A. В. Черкасов, Методи управління екологічними ризиками в системі забезпечення економічного розвитку регіону, 2011

A. В. Щербак & М. В. Зеленська & Ю. М. Гайдай, Льодовий режим річок України (льодоутворення та його характеристики), 2011

Щоб знайти усі документи, що створені у 2011 і починаються з займенника, запит оформлюється так:

<doc date="2011"> [tag=".*pron.*"]

ксті вибору її інтеграційного вектора, 2011	задач , які виникають в науках про Землю . </doc><doc> Однею з головних характеристик сучасного
изнаних держав, 2011	компанія « Інтобуд » . Злочин не розкрито . </doc><doc> ВСЕ , що є СТАЄ МИНУЛИМ Певен , їх і з
ла Старицького Богдан Хмельницький"" , 2011	завершальному етапі Другої світової війни . </doc><doc> ДЕЯКИ ПРОБЛЕМИ МІЖНАРОДНО-ПРАВОВОГО
х досліджень проблематики політичної волі, 2011	яке вчинене суб'єктом і самого суб'єкта . </doc><doc> ОДНА З ДУМНИХ РЕМІНІСЦЕНЦІЙ У ТРИЛОГІ
рської діяльності Михайла Булгакова, 2011	працівників як носіїв людського фактору . </doc><doc> ДЕЯКИ ОСОБЛИВОСТІ ФІЛОСОФСЬКИХ ТА КФ
	цінностями , домінуючими у певному суспільстві . </doc><doc> ДЕЯКИ ДОКУМЕНТИ З ФОНДІВ ЦДІАК УКРАЇН
	всі ! До наступного тижня ! До зустрічі ! </doc><doc> Мойй сім'ї , в якій любов поєднала українці
рдо Сандрі, 2011	а ще проєкт мандрівних кіноконцертів . </doc><doc> ЧОМУ ВСЕ ЛИШЕ ОБРУШУЄТЬСЯ ?! В україн
юлітичних процесів, 2011	. Фрагменти Запитувала Катерина Щоткіна </doc><doc> Його поведінка , його спосіб реагування в
онання мирової угоди, 2011	суспільної або особистої потреби [2] . </doc><doc> ДЕЯКИ ТЕОРЕТИЧНІ ЗАСАДИ МОДЕЛЬНОГО Е
раїни), 2011	й з якості корпоративного управління . </doc><doc> ДЕЯКИ ПИТАННЯ ПРИМУСОВОГО ВИКОНАНН
нтуційного права, 2011	суспільства та Української держави загалом . </doc><doc> Вона була першою Розкрито плідну діяльн
безпеки України процедур, що обмежують права і свободи людини і громадянина, 2011	руках тогочасного політичного режиму . </doc><doc> ДЕЯКИ АСПЕКТИ РОЗУМІННЯ НАЦІЇ ЯК СУБ'ЄІ
	кризи , яку сьогодні переживає Україна . </doc><doc> ДЕЯКИ АСПЕКТИ ЗАСТОСУВАННЯ СЛУЖБОО
зму судочинства, 2011	стимулює загального розвитку виробництва . </doc><doc> Всі історики певним чином перебувають
перативно-розшукової діяльності у кримінальному процесі, 2011	у найближчому майбутньому не уникнути ... </doc><doc> ДЕЯКИ АСПЕКТИ ЗАБЕЗПЕЧЕННЯ ПОЗОВУ В Г
тарил)-2-імінокумаринів з алкілюючими агентами, 2011	Уніфікованим правилам по інкасо тощо) . </doc><doc> ДЕЯКИ ОСОБЛИВОСТІ ВИКОРИСТАННЯ РЕЗУ,
іоритетів міграційної політики України, 2011	відносяться до бюджетного законодавства . </doc><doc> Як відомо , різноманітні похідні , особл
фері боротьби з відмиванням «брудних» грошей, 2011	задекларованим підгрунтам державотворення . </doc><doc> ДЕЯКИ АСПЕКТИ ВИЗНАЧЕННЯ КОНЦЕПТУАЛ
ки?, 2011	митної справи на митному кордоні України . </doc><doc> ДЕЯКИ АСПЕКТИ ВДОСКОНАЛЕННЯ ЗАКОНОД
	продовжуватиме вважатися абсолютною цінністю . </doc><doc> ЯК ЗМІНЮЄТЬСЯ ЛЕКСИКОН СУЧАСНОЇ У
	тітка Файна побігла на буряки ... 2011 р. </doc><doc> Колись , у третьому класі , сусід по парті Ва
	відгукнулася на його регіт своїм ... Квітень 2011 </doc><doc> Тепер дівчата приїздили вранішньою марш
ізичних досліджень будови Приазовського мегаблока Українського щита, 2011	затоплених шахт та карстових порожнин . </doc><doc> ДЕЯКИ РЕЗУЛЬТАТИ КОМПЛЕКСНИХ ГЕОЛОГ
аїні), 2011	авіаційного гасу в геологічному середовищі . </doc><doc> ДЕЯКИ ПРОБЛЕМИ ОХОРОНИ ВОДНИХ РЕСУР
я державної етнонаціональної політики в Україні, 2011	нормативно-правовий інформаційний ресурс . </doc><doc> ДЕЯКИ НАУКОВІ ПІДХОДИ ДО ВИЗНАЧЕННЯ І
римінальному процесі, 2011	підхід при розгляді дисертаційних робіт . </doc><doc> ДЕЯКИ ГАРАНТІЇ ОХОРОНИ АДВОКАТСЬКОЇ Т
ювну норму в освітній літературі), 2011	лексикси від найдавніших часів до сьогодення . </doc><doc> Як вчимося сьогодні , так житимемо за
оже рідною мовою! (Листи П.С. Морачевського до І.І. Срезневського 1862, 1864, 1874 рр.), 2011	варто плутати із звичайною неграмотністю . </doc><doc> ЩО МОЖЕ БУТИ СВЯТІШЕ ДЛЯ СЕРЦЯ НА

Щоб знайти всі форми дієслова *бути* всередині документів, позначених як тексти художнього стилю, слід написати запит:

[lemma="бути"] within <doc style="FIC"/>

Григорій Косинка, Мати, 1925	це далекий наш родич : коли що трапилось дома , тут	буде	відомо . Але я не розумію його хвилювання . Він одчинив
Григорій Косинка, Мати, 1925	пристаркуватий з великим шрамом на бороді вид такий	був	збитий за життя , що я мимоволі згадав його слова
Григорій Косинка, Мати, 1925	... Хліба трудніше достать , а вода , слава богу ,	є	! Не вірите в бога ? Не годиться так ... Падають
Григорій Косинка, Мати, 1925	діти , хоч раз десь ізнятися з матір'ю , щоб пам'ять	була	, — той самий голос старої . Я виходжу з хати ; принесли
Григорій Косинка, Мати, 1925	Але в мене справді не знайшлося слова , яким можна	було	б на хресті матері змалювати її страждання . </doc>
Григорій Косинка, В житах, 1925	змалювати її страждання . </doc><doc> В житах Це все	було	просто до дрібниць : і я , і заспаний ранок , і сивий
Григорій Косинка, В житах, 1925	раз , а друге — в нашій революційній термінології це	є	дезертир , а я , товариші , саме до них і належав
Григорій Косинка, В житах, 1925	і твердо рішає : " Убити двох , трьох , а тоді що	буде	... Застрелиться " . Але мимоволі кладу за обніжок
Григорій Косинка, В житах, 1925	подавай !.. А другий на возі : — Комісарами хотять	бути	. Комісарами ?! Хай чортом буде ! А то як ніч — з
Григорій Косинка, В житах, 1925	Комісарами хотять бути . Комісарами ?! Хай чортом	буде	! А то як ніч — з винтовкою до вікна йде : " Дайош
Григорій Косинка, В житах, 1925	медовнику бджола , посміхаюсь та лізу в густий льон . Хай	буде	й так ... Дзінь , Дзюба , дзінь ... Це дзвонить степ
Григорій Косинка, В житах, 1925	Подумаєш , герой який знайшовся : " Комісарами хотять	бути	" ? А хоч би й комісарами ?.. Ні , Гострому цього
Григорій Косинка, В житах, 1925	большовик , А я — гайдамачка ! " І підморгне ! Молодець	був	Матвій , коли вели його ... Дзінь ... Про комісарів я
Григорій Косинка, В житах, 1925	" Я од здивування підсунув на лоба картуз : " Що ж	буде	дали ? " Стояла переді мною справжня Уляна , а з
Григорій Косинка, В житах, 1925	... Воду носила . — Здарстуй ! — і стала . — Здорова	будь	, Уляно ! — хотів посміхнутись і не міг : вона довго
Григорій Косинка, В житах, 1925	Корнію ... Хочеш цілувати ? Цілуй , хай хоч один день	буде	наш ! І гладила рукою мій чуб , а його розчісували
Григорій Косинка, В житах, 1925	Дзюби ! Я поклав голову на її коліна і слухав , бо це	була	загублена в житах моя доля : — У мене так наче пісню
Григорій Косинка, В житах, 1925	мережку , на пазусі — кленове листя , і все кругом	було	п'яне , а червона хустка зайнялася й горіла степом
Григорій Косинка, В житах, 1925	— Тихо поцілувала , рвонула льону горстку , й очі	були	сині-сині , мов льон , а хустка гасла . — Прощай
Гук, По дорози додому, 1925	й безпорядки повелись по залізницях ! Раніш треба	було	тобі сісти у вагона — трах корзиною у вікно , повкидав
Гук, По дорози додому, 1925	а сам думаю : — Чи в той же я потяг сів ? От добре	буде	, як замість Кобеляк та завезуть в другий бік . Спитаю
Гук, По дорози додому, 1925	замість Кобеляк та завезуть в другий бік . Спитаю .	Будьте	ласкаві , ви куди їдете ? Кременчук . Так . Ну ,

Такі самі умови для документів можна задавати й за допомогою користувацького інтерфейсу на сторінці пошуку:

DOC.STYLE

- ACA
- EGO
- FIC
- FOL
- JOU
- OFF
- POE
- REL
- SPO

Вибрати все

Групування

Деякі операції доречно застосовувати не до одного символу/токена, а до групи. Щоб кілька символів або токенів опрацьовувалися

як цілісна група, вони беруться у круглі дужки. Приміром, щоб шукати 2 дієслівно-іменні групи підряд, можна застосувати запит

```
запит ([tag=".*verb.*"][tag=".*noun.*"]){2}
```

Результат запиту: "має намір взяти участь", "змусили противника припинити обстріл" тощо

Між групами можна використовувати логічне "АБО", що позначається вертикальною рисою - | , наприклад:

```
([tag=".*past.*"][lemma="би|б"])([lemma="би|б"][tag=".*past.*"])
```

В результаті запиту будуть знайдені такі приклади "ходив би", "би ходив" тощо.

Логіка групування застосовна як до мови CQL, так і до мови регулярних виразів.

Групування - це потужний інструмент для пошуку токенів з деякими комбінаціями без обов'язкового вказання символів, що їх складають.

1. Щоб знайти слова з однаковими сполученнями літер у будь-якій частині слова, де сполучення складається з 5 літер, можна скористатися таким запитом до ГРАКу:

```
[word="(\\w{5}).*\\1.*"]
```

- `(\\w{5})` - це перша і єдина група захоплення в цьому запиті. Вона шукає послідовність рівно з п'яти символів. Дужки () створюють групу, яку можна потім використати повторно.
- `.*` - це послідовність будь-яких символів (включаючи відсутність символів) між першим входженням групи та її повторенням.
- `\\1` - це зворотне посилання на першу групу захоплення. Воно шукає точне повторення тих самих п'яти символів, які були знайдені на початку.
- `.*` - це ще одна послідовність будь-яких символів після повторення до кінця слова.

Приклад знайдених слів: "ледве-ледве", "зовсім-зовсім", "видимо-невидимо". Важливо враховувати, що групи захоплення діють у межах того самого токена (тобто, приміром, групу номер один з першого токена не можна застосувати до іншого токена - у разі потреби в пошуку повторів за межами одного токена у CQL слід застосовувати мітки - див. "Використання глобальних умов для порівняння токенів, пошуку однакових словоформ і лексем").

2. Для пошуку слів (наприклад, вигуків) з послідовними повторами будь-якої кількості можна використати:

```
[word="(w+)(?:\W*\1)+.*"&tag=".*intj.*"]
```

Цей запит знайде вигуки з повторюваними частинами, як-от "тю-тю-тю-тю", "ги-ги-ги", "а-а-а".

1. $(w+)$ - це перша група захоплення. Вона шукає одну або більше літер (або цифр, або підкреслення). Ця група "запам'ятовується" для подальшого використання.
2. $(?:\W*\1)+$ - це незахоплююча група, яка повторюється один або більше разів:
 - $(?:\dots)$ - незахоплююча група, яка не створює окремого посилання
 - $\W*$ - нуль або більше нелітерних символів (наприклад, дефіс)
 - $\1$ - зворотне посилання на першу групу, тобто повторення раніше знайденого сполучення літер
3. $.*$ - будь-яка кількість будь-яких символів до кінця слова

3. Щоб знайти слова з повторами на початку та в кінці, використовуємо:

```
[word="^(w{2,})w+\1$"]
```

Приклад слів: "заліза", "можемо", "намистинами".

1. $^$ - вказує на початок слова.
2. $(w\{2, \})$ - це група захоплення, яка шукає два або більше словесних символів (літер, цифр або підкреслення) на початку слова. $\{2, \}$ означає "два або більше".
3. $w+$ - це послідовність одного або більше словесних символів, що йдуть після початкового сполучення.
4. $\1$ - зворотне посилання на першу групу захоплення, тобто повторення того самого сполучення літер, яке було на початку слова.
5. $\$$ - вказує на кінець слова.

4. Для пошуку слів з двома послідовними сполученнями, які потім повторюються, можна застосувати :

```
[word="(w{4})(w{4}).*\1\2.*"]
```

Цей запит знайде слова на кшталт "вантаження-розвантаження", "асимільованих-переасимільованих", "сімнадцяти-вісімнадцяти", "карбонатно-бікарбонатному", "Перемишль-Львів-Перемишль".

- $(w\{4 \})$ - перша група захоплення, яка шукає рівно чотири символи.

- $(\backslash w\{4})$ - друга група захоплення, яка також шукає рівно чотири словесних символи, одразу після першої групи.
- $.*$ - дозволяє будь-яку кількість будь-яких символів (або жодного) між початковими групами та їх повторенням.
- $\backslash 1$ - зворотне посилання на першу групу захоплення, шукає точне повторення перших чотирьох символів.
- $\backslash 2$ - зворотне посилання на другу групу захоплення, шукає точне повторення других чотирьох символів.
- $.*$ - дозволяє будь-яку кількість будь-яких символів (або жодного) після повторення до кінця слова.

Усі ці запити використовують групування $(\backslash w+)$, щоб "запам'ятати" певне сполучення літер, і потім шукають його повторення за допомогою зворотних посилань (згадані $\backslash 1$ та $\backslash 2$).

Кількість груп можна збільшувати та задавати їх склад, наприклад:

```
[word="(при|за|на)(біг|літ|хід)(\w{2,3})(ся)?" & tag="verb.*"]
```

1. (при|за|на) - перша група, яка шукає один з трьох префіксів.
2. (біг|літ|хід) - друга група, яка шукає один з трьох коренів.
3. $(\backslash w\{2,3})$ - третя група, яка захоплює 2-3 літери після кореня (суфікс).
4. $(\text{ся})?$ - необов'язкова четверта група, яка шукає "ся".

Цей запит знайде такі слова: "залітали", "прибігав", "набігався".

Підходи можна комбінувати та таким чином знаходити майже будь-які слова в корпусі, зазначивши в запиті лише частину бажаного слова.

Глосарій вжитих термінів

KWIC див. **Ключові слова**

Абсолютна частота – див. **Частотність**.

Атрибут токена – параметр токена, за яким ведеться пошук; атрибут може вказувати на лексему, тег, словоформу (у запиті – lemma, tag, word).

Атрибут документа – інформація, що додається до документа в корпусі, називається метаданими. Наприклад, doc.style вказує на стиль документа, doc.author – на автора, doc.sex — на стать автора.

Документ – загальна назва, яка використовується для позначення кожного з текстових файлів у складі корпусу. Початок і кінець кожного документа в корпусі автоматично позначаються як `<doc>` і `</doc>`.

Запит – послідовність символів, що вводить користувач для пошуку в корпусі з метою отримання результатів у вигляді конкордансу.

Запит CQL — запит, написаний мовою CQL, який описує характеристики токена або послідовності токенів і, можливо, якісь додаткові умови для них.

Ключові слова – це слова, за якими здійснювався пошук, вони виділені червоним у конкордансі. KWIC – аббревіатура, що означає ключове слово в контексті, Key Word in Context.

Конкорданс – перелік усіх прикладів пошукового слова або фрази, знайдених у корпусі, з інформацією про джерело (метаданими). Можна налаштувати вигляд конкордансу: із пошуковим словом, виділеним у центрі екрана, та з контекстом праворуч і ліворуч, (формат KWIC) або із пошуковим словом, виділеним кольором у реченні (формат «У реченні»).

Корпус – велике зібрання текстів, призначене для вивчення мови, що репрезентує зріз мови в цілому або окрему її частину. Корпус зазвичай буває анотований, або розмічений, тобто містить певну інформацію про тексти (наприклад, стиль, жанр, рік написання) і словоформи (граматичні, семантичні, стилістичні характеристики). Терміни «корпус», «корпус текстів» і «мовний корпус» — взаємозамінні. Використання корпусу для будь-якого типу лінгвістичної чи мовно-орієнтованої роботи забезпечує результати, що відображають реальне використання мови.

Лема – початкова форма слова, позиційний атрибут, який програма приписує кожній словоформі в корпусі за словником (отже, леми визначаються тільки для словоформ, які є в словнику, а також для форм складних слів, написаних через дефіс, кожна частина яких є у словнику як окрема лема, інші залишаються без лем). Лематизований корпус дозволяє здійснювати пошук лемою і включати в результат усі форми слова, наприклад, шукаючи лемою *ити*, знайдемо *іду, іде, ідемо*. Лема чутлива до реєстру, тобто *лев* і *Лев* — дві різні леми. Частотний список за лемою — це список частот, де всі *іду, іде, ідемо* подано як *ити*. Різні види дієслова, орфоепічні і правописні варіанти (на зразок *і/ї, у/в, вчитель/учитель, аудиторія/авдиторія*) у ГРАКу вважаються різними лемами. Пошук за лемою дає змогу користувачеві отримати дані, що точніше характеризують лексему в цілому (її частотні характеристики, сполучуваність з іншими лексемами у рамках словосполучень тощо). Тому цей атрибут слід використовувати в усіх запитах, спрямованих на виявлення характеристик саме лексеми, а не її окремої граматичної форми. Винятками є ненормативні, помилкові, обценні, новостворені, архаїчні лексеми, яких немає в словнику — їх можна знайти тільки за словоформою (наприклад, [word="модняр.*"]).

Метадані — дані з характеристиками певного документа (тексту); наприклад, дані про стиль, автора, стать автора, рік створення, рік публікації тексту.

Позиційний атрибут – інформація, що стосується токена в корпусі, наприклад, словоформа (word), лема (lemma), морфологічні ознаки (tag).

Словоформа – реалізація лексеми в тексті, може бути в будь-якій граматичній формі (наприклад, *іду, іде, ідемо* для леми *ити*); під час пошуку токенів за словоформою використовується позиційний атрибут “word”. Розділові знаки, цифри, будь-які неалфавітні символи й слова іншими мовами також можна шукати за атрибутом “word”.

Структура – корпусна структура це один з автоматично маркованих сегментів, на які можна поділити корпус. ГРАК поділяється на документи (<doc> </doc>) та речення (<s> </s>).

Тег – мітка, присвоєна токенам в анотованому корпусі для позначення граматичних, семантичних та стилістичних характеристик. Наприклад, у наборі тегів для словоформи *Ганна* – |noun:anim:f:v_naz:prop:fname| граматичними характеристиками є

poun:anim:f:v_naz: (іменник, істота, жіночий рід, називний відмінок), семантичними – rgor:fname (власна назва, ім'я). Приклади стилістичних тегів: coll – розмовне, slang – сленгове, rare – рідковживане, діалектне або застаріле.

Токен – найменша одиниця, на яку поділяється корпус. Такою найменшою одиницею є будь-яка послідовність символів між пробілами або іншими роздільниками: словоформа, число, розділовий знак, символ (смайлик, математичний символ тощо). Розділові знаки в корпусі опрацьовуються як окремі токени. Пробіли — це не токени.

Частотний список – список лем, словоформ або тегів, побудований за певним корпусом, частиною корпусу, конкордансом або вибіркою, де для кожного елемента списку вказано його частотність.

Частотність (також **абсолютна частотність**) – кількість випадків появи слова, фрази, тега тощо в корпусі, підкорпусі, конкордансі або вибірці. Якщо слово має частоту 10, це означає, що воно було знайдено 10 разів. Це абсолютна цифра, яка не розраховується за допомогою конкретної формули.

Переклад документації SketchEngine: Анна Власенко, Олена Гульцьо, Анастасія Худяк, Юлія Лісовська, Юлія Гадяк, Олена Кулява

Редагування, адаптація: Марія Шведова, Сергій Фокін, Арсеній Лукашевський.