
VivA se Korpusportaal gebruikershandleiding

Weergawe 1.0 - Mei 2017

Liesbeth Augustinus



09/05/2017



VIRTUELE INSTITUUT VIR AFRIKAANS

Inhoudsopgawe

1	Wat is korpuslinguistiek?.....	1
2	Wat is Korpusportaal?.....	1
3	Korpusinligting	2
3.1	Korpus	3
3.2	Statistiek	3
3.3	N-gramme.....	4
4	Soek	5
4.1	Eenvoudig.....	5
4.1.1	Soekopdrag	5
4.1.2	Resultate	5
4.2	Uitgebrei	9
4.2.1	Soekopdrag	9
4.2.2	Resultate	10
4.2.3	Bondelnavrae	10
4.3	Gevorderd.....	12
4.3.1	Soekopdrag	12
4.3.2	Resultate	13
4.3.3	Bondelnavrae	14
4.4	Kundig	14
4.4.1	Soekopdrag	14
4.4.2	Resultate	15
5	Corpus Query Language (CQL)	16
5.1	Eenvoudige soekopdragte	16
5.1.1	Enkele basiese beginsels	16
5.1.2	Soek na 'n woord, lemma of woordsoort	16
5.1.3	Soek na naasliggende woorde, lemmas en/of woordsoorte	17
5.1.4	Soek na nienaasliggende konstruksies.....	17
5.2	Boolese operatore.....	18
5.2.1	AND	18
5.2.2	OR.....	18
5.2.3	NOT	18
5.3	Vaste uitdrukings.....	19
5.3.1	Oorheersstring vir 'n willekeurige karakter	19
5.3.2	Groepering van karakters en patrone.....	19



5.3.3	Kwantifikasie	19
5.3.4	Ontsnaptteken.....	20
5.3.5	Kassensitiewe soekopdragte.....	20
6	Addendum	21
6.1	Korpora in Korpusportaal	21
6.2	Linguistiese annotasie	21
6.2.1	Etikette	21
6.2.2	Woordsoorte algemeen	22
6.2.3	Woordsoorte gedetailleerd.....	22
6.2.4	Boolese operatore.....	26
6.2.5	Regex-operatore	26
6.2.6	Kwantifikasie van soektogomgewing.....	27

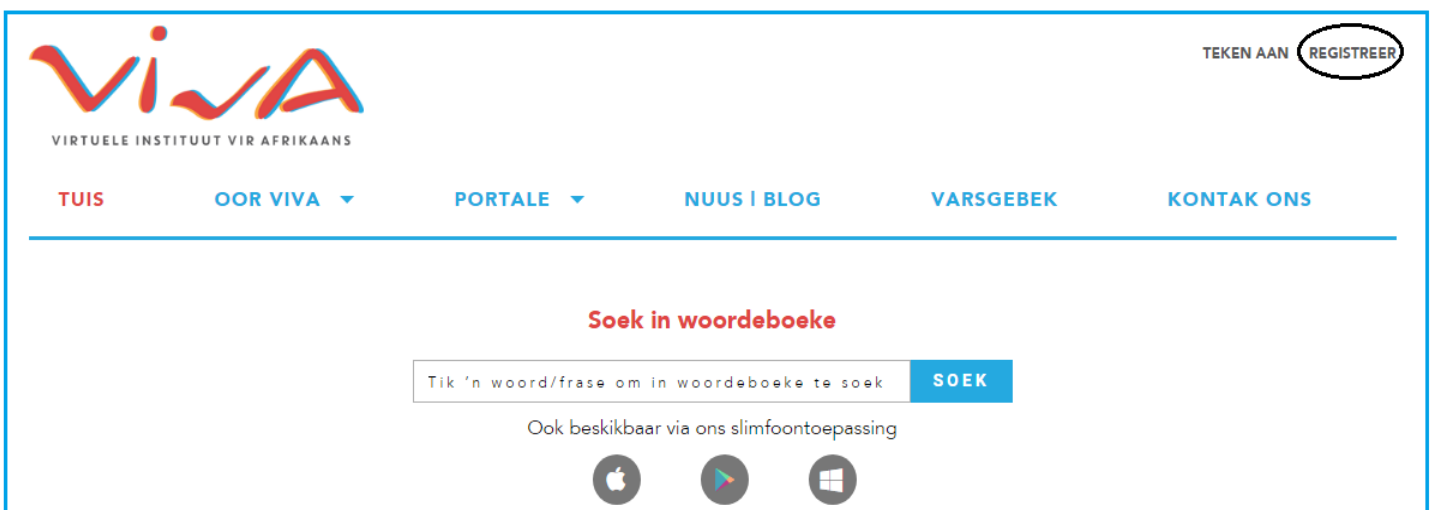
1 Wat is korpuslinguistiek?

Binne die domein van die taalwetenskap is daar verskillende ondersoekmetodes. Etlike dekades reeds baan introspektiewe benaderinge die weg vir meer empiries gerigte ondersoektegnieke, wat eerder berus op werklike taalgebruik as op die grammatikale oordeel van die taalkundige self. Korpuslinguistiek vorm deel van die kwantitatiewe empiriese benaderinge, waardeur taalkundige verskynsels met behulp van groot databasisse van gesproke of geskrewe taal, d.w.s. korpora (enkelvoud: korpus), ondersoek word.

2 Wat is Korpusportaal?

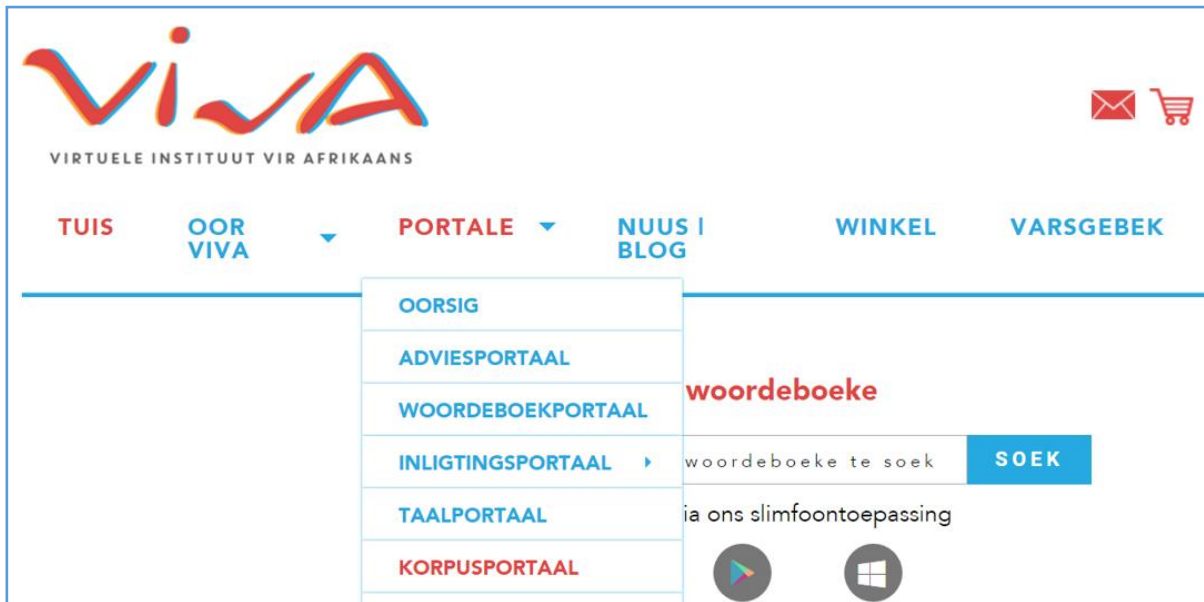
Binne die bestek van die Viva-projek (<http://viva-afrikaans.org>) is Korpusportaal sedert Oktober 2015 vir Afrikaans toeganklik gemaak, waardeur meer as **85 miljoen woorde** in geskrewe Afrikaans ontsluit is. Die Korpusportaal bevat naas die korpora self ook 'n **uitgebreide soekfunksie** wat die moontlikheid bied om die data te deursoek.

Om toegang te verkry tot die Korpusportaal moet gebruikers eers **registreer** en 'n jaarlikse subskripsie van R100 betaal. Voornemende gebruikers kan registreer deur aan die regterkant van die landingsblad op "registreer" te klik.

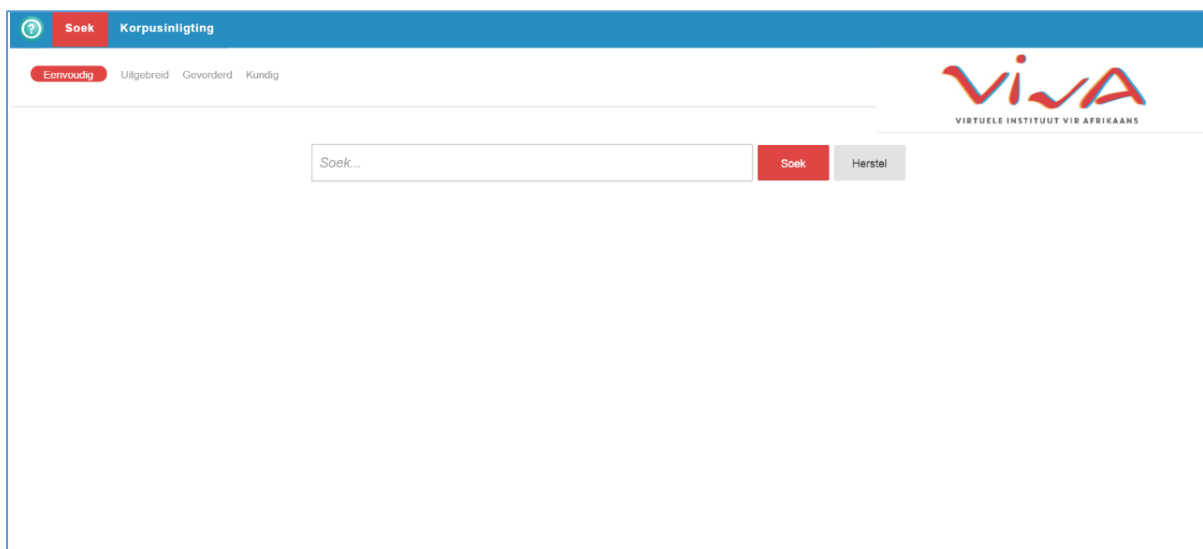


The screenshot shows the homepage of the Virtuele Instituut vir Afrikaans (Viva). The logo 'viva' is in red and blue, with 'VIRTUELE INSTITUUT VIR AFRIKAANS' below it. In the top right corner, there is a 'TEKEN AAN' button and a circled 'REGISTREER' button. The navigation menu includes 'TUIS', 'OOR VIVA', 'PORTALE', 'NUUS | BLOG', 'VARSGEBEK', and 'KONTAK ONS'. Below the navigation is a search section titled 'Soek in woordeboeke' with a search input field containing the placeholder text 'Tik 'n woord/frase om in woordeboeke te soek' and a blue 'SOEK' button. Below the search field, it says 'Ook beskikbaar via ons slimfoontoepassing' and shows icons for Apple, Google Play, and Windows.


Sodra registrasie afgehandel is, kan jy aanteken deur op die skakel te klik. Daarna kan jy in die kieslys bo oor "Portale" huiwer, waarna jy op "Korpusportaal" kan klik wat jou na die regte blad sal neem (<http://korpus.viva-afrikaans.org/page/search>).



Die Korpusportaalblad lyk so:



Die Korpusportaalblad het 2 opsies bo: “Soek” en “Korpusinligting”. Afdeling 3 beskryf watter inligting onder "Korpusinligting" te vinde is. Die soekoppelvlak word in afdeling 4 bespreek.

Met die -knoppie kan jy 'n stelselmatige verkenningstog van die Korpusportaal onderneem, waarin elke funksionaliteit uiteengesit word.

3 Korpusinligting

In die Korpusportaal is daar op die “Korpusinligting”-blad drie opsies met inligting oor die samestelling van die korpora: Korpus, Statistiek en N-gramme.

3.1 Korpus

Op die “Korpus”-blad is daar ’n oorsig oor die verskillende korpora en hulle spesifikasies, waaronder genre en omvang (aantal woorde) van die tekste (sien ook Addendum A). Alles ingereken bevat die Korpusportaal meer as 85 miljoen woorde in geskrewe Afrikaans. Die meeste brontekste is baie reser, en kom uit die tydperk 1995-2015.

Alle korpora is outomaties gelemmatiseer (omtrent 90% akkuraat) en van woordsoorte voorsien (omtrent 75% akkuraat). Die annotasies is nie gekorrigeer nie; gebruikers moet dus rekening hou met moontlike annotasiefoute.

Soek **Korpusinligting**

Korpus Statistiek N-gramme

VIVA
VIRTUELE INSTITUUT VIR AFRIKAANS

Gebruiksvoorwaardes
Alle korpora is uitsluitlik in Viva se Korpusportaal vir navorsingsdoeleindes beskikbaar. [Kontak Viva](#) vir navrae oor ander lisensieopsies.

Annotasie
Alle korpora is outomaties gelemmatiseer (~90% akkuraat) en van woordsoortetikette voorsien (~75% akkuraat). Geen kwaliteitskontrole toegepas nie.

NCHLT-Afrikaanse korpus 1.0
’n Versameling regeringsdokumente van die onderskeie webblaaie van die Suid-Afrikaanse regeringswebblaaie en elektroniese publikasies.

Verwysing
Department Kuns en Kultuur & CText. 2013. Afrikaans NCHLT Text Corpora. Potchefstroom: CText, North-West University. ISLRN: 544-932-849-161-3. Available: <http://rma.nwu.ac.za/index.php/resource-catalogue/afrikaans-nchlt-text-corpora.html>

Grootte
3,200,000 woorde

NWU/Maroela Media-korpus 1.0
Versameling webartikels soos gepubliseer op die Maroela Media-webtuiste.

Verwysing
Maroela Media en CText. 2015. NWU-Maroele Media-korpus 1.0. Potchefstroom: CText, Noordwes-Universiteit.

3.2 Statistiek

Onder “Statistiek” kan jy vir elke outeur, publikasiedatum en korpuskomponent (d.w.s. metadatafilters) **frekwensierangskikkings** opvra vir woorde, lemmas en woordsoorte. Deur op die plusteken regs te klik, kan jy ekstra metadatafilters skep; deur op die minus te klik, kan jy dit weer verwyder. Die figuur hieronder illustreer die soekopdrag om die mees frekwente woord in die RSG-nuuskorpus vir 2014 op te roep. Die “stoor”-knoppie stel jou in staat om die resultaat uit te voer. Dit is verder moontlik om ’n dokumentlys aan te vra, ’n grafiek van die frekwensie van die inskrywing (“woordeskattgroei”) aan te vra en ’n woordwolk te genereer.

Soek Korpusinligting

Korpus **Statistiek** N-gramme

Metadatafilters

Pas hierdie reëls toe:

Versamelingnaam is RSG-nuuskorpus
 Publikasiedatum is 2014

Frekwensielystipe woord

Soek Herstel

Filter CollectionName="RSG-nuuskorpus"&PublicationDate="2014"
 Frekwensielystipe word
 Unieke trefslae 53392
 Trefslae 1786143
 Dokumente 1

Frekwensielys **Dokumentys** Woordeskatgroei Woordwolk

Toon 50 per bladsy

<< < 1 2 3 4 5 6 ... > >>

Bladsy 1 van 1068

Gaan

Stoor

Groep ▲ ▼	Trefslae ▲ ▼
die	71179
.	48728
.	28153
in	28001
het	27843

3.3 N-gramme

By “N-gramme” kan jy ’n frekwensielys van N-gramme genereer. ’n N-gram verwys binne hierdie konteks na ’n opeenvolging van N woorde, lemmas of woordsoorte (of ’n kombinasie daarvan), met ’n maksimumlengte van 5. Die soekopdrag hieronder lewer derhalwe ’n antwoord op die vraag watter substantiewe die meeste voorkom ná die voorsetsel of in die werk van Cecilia Nortjé.

Metadatafilters

Pas hierdie reëls toe:

Outeursnaam / Skulnaam is CECILIA NORTJÉ

N-gramgrootte 2 N-gramtipe woord

woord	woordsoort	woord	woord	woord
op	Selfstandige naamwoord	<enige>	<enige>	<enige>

Soek Herstel

Query [word="op"][pos="S.NW.*"]
 Filter AuthorNameOrPseudonym="CECILIA NORTJÉ"
 N-gramgrootte 2
 N-gramtipe woord
 Aantal n-gramme 19
 Trefslae 51
 Dokumente 2

Toon 50 per bladsy

Bladsy 1 van 1 Gaan

Voer uit

Groep ▲ ▼	Trefslae ▲ ▼
op pad	18
op aarde	9
op Spear	3
Op pad	2
op horings	2

4 Soek

Die soekoppelvlak (“Soek”) bied vier verskillende maniere om in die korpora te soek: “eenvoudig”, “uitgebreid”, “gevorderd” en “kundig”.

4.1 Eenvoudig

4.1.1 Soekopdrag

Met die opsie “eenvoudig” kan jy via ’n soekblokkie een woord of ’n opeenvolging van woorde in die volledige Korpusportaal opsoek.¹ Daar is nie ’n opsie om metadata te filter nie. Daarbenewens is die soekterm nie kassensitief nie, wat beteken dat onder meer hoofletters en aksenttekens by die soektog ingesluit word.

Soek Korpusinligting

Eenvoudig Uitgebreid Gevorderd Kundig

baie maklik

Soek Herstel

VIVA
VIRTUELE INSTITUUT VIR AFRIKAANS

4.1.2 Resultate

Nadat jy op “soek” geklik het, word die soekterm in alle korpora van die Korpusportaal gesoek. Hieronder word ’n aantal trefslae vir die soekopdrag “baie maklik” vertoon.

¹ Dit is ook moontlik om ’n vaste uitdrukking te gebruik, kyk afdeling 5.3.

Soek Korpusingligting

Eenvoudig Uitgebreid Gevorderd Kundig || Resultate

Soekopdragte:

#	Soekopdrag	binne	Metadatafilters	Groepering	Status	Trefslae	Dokumente	
2	[(word="baie")][(word="maklik")]	document		-	KLAAR	359	66	VERANDER X

Soekresultate:

Trefslae | **Dokumente** | Trefslae in groepe | Dokumente in groepe

Toon 50 per bladsy << < 1 2 3 4 5 6 ... > >> Toggle titels Bladsy 1 van 8 Gaan

Stoor resultate

Konteks links	Trefstuk ▲ ▼	Konteks regs	lemma ▲ ▼	woordsoort ▲ ▼
... nie , het die spankeuse	baie maklik	gemaak . Die Australiese afrigter ...	baie maklik	TW.hoofdelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... droog met lang gras wat	baie maklik	aan die brand kan slaan ...	baie maklik	TW.hoofdelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... eie bedryf , omdat dit	baie maklik	is om goedkoop , maar ...	baie maklik	TW.hoofdelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... die lig gebring dat dit	baie maklik	is om 'n selfoon se ...	baie maklik	TW.hoofdelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... Volgens ons verslaggewer het hy	baie maklik	op Parkstasie in Johannesburg geholpe ...	baie maklik	TW.hoofdelwoord.adjektief.onbepaald B.NW.stellend.predikatief
... minuut af aangeval en dit	baie maklik	laat lyk . Die tuisspan ...	baie maklik	TW.hoofdelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... maak , en inmenging kan	baie maklik	vir die doeleindes misbruik word ...	baie maklik	TW.hoofdelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... wat toegang tot die skool	baie maklik	maak . Die voorsitter van ...	baie maklik	TW.hoofdelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... wat toegang tot die skool	baie maklik	maak . Die voorsitter van ...	baie maklik	TW.hoofdelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... Potgieter , sê dit is	baie maklik	vir misdadigers om die toestelle ...	baie maklik	TW.hoofdelwoord.bywoord.onbepaald B.NW.stellend.predikatief

Bo aan die resultateblad ("resultate") word weergegee: die soekopdrag in Corpus Query Language (kyk afdeling 5), die status (is die soekopdrag afgehandel of nie), die aantal trefslae en die aantal dokumente waarin die trefslae voorgekom het. Jy kan die soekopdrag aanpas deur op "verander" te klik.

Waarskuwing: As 'n soekopdrag meer as 5 miljoen trefslae oplewer, sal daar onder "trefslae" 5000000 weergegee word. Die syfer moet dus geïnterpreteer word as "5 miljoen of meer".

Onder die soekopdrag is die venster met soekresultate. Onder die opskrif "**Trefslae**" word die resultate gegee as keywords in konteks (KWIC).

Die woorde wat met die soekopdrag ooreenkom (d.w.s. die trefslae of die "trefstuk") is vetgedruk en word op dié manier visueel van die linker- en regterkonteks geskei. Daarbenewens word die lemmas en woordsoorte van die trefslae gegee. Die resultate kan gesorteer word deur op een van die kolomname te klik. Jy kan kies of jy 50, 100 of 200 resultate per bladsy te sien wil kry. Deur op "Toggle titles" te klik, kan jy inligting sien oor die bron van die sin, d.w.s. die outeur en die publikasiedatum van die teks. Deur op die soekresultaat te klik, kan jy meer konteks sien. Jy kan die resultate stoor deur op "stoor resultate" te klik.

Onder die opskrif "**Dokumente**" word die trefslae gesorteer volgens die teks ("dokumenttitel") en korpus ("versameling"). So kan jy byvoorbeeld vasstel watter tekste die meeste trefslae oplewer.

Soek Korpusinligting

Eenvoudig Uitgebreid Gevorderd Kuldig || Resultate Dokument

Soekopdragte:

#	Soekopdrag	binne	Metadatafilters	Groepering	Status	Trefslae	Dokumente	
1	[[word="baie"]][[word="maklik"]]	document			KLAAR	359	66	VERANDER

Soekresultate:

Trefslae Dokumente Trefslae in groepe Dokumente in groepe

Toon 50 per bladsy << < 1 2 > >> Bladsy 1 van 2 Gaar

Stoor resultate

Dokumenttitel ▼	Versameling ▼	Hits ▲
147. Nie-fiksie Nie-akademies Koerant Landswyd Rapport deur Onbekend ... nie , kan die mark baie maklik die vlak van 28 000 kan verbeter . ' Dis baie maklik vir my om in ? , maak die tegnologie dit baie maklik vir vooraanstaande universiteite om hoër het die ondersoek vir ons baie maklik gemaak . Selfs Oprah het ontdek dat sindikate haar geld baie maklik elektronies kan rondskuif . 'n ...	Taalkommissie-korpus.Nie-fiksie.Nie-akademies.Koerant.Landswyd.Rapport	50
148. Nie-fiksie Nie-akademies Koerant Suid Burger deur Onbekend ... is baie vulnerable en raak baie maklik by verkeerde dinge en wiendekringe al dan nie , kan baie maklik vir Kaapstad deelname aan die aanvalsbewegings is goed gedoen en baie maklik om baas te raak nie . Die vinger word baie maklik na banke gewys . Maar misluk . Die vinger word baie maklik na banke gewys . Maar ...	Taalkommissie-korpus.Nie-fiksie.Nie-akademies.Koerant.Suid.Burger	48
127. Nie-fiksie Akademies Studiegids VerskeieWetenskappe3 deur Onbekend ... verwys , dit vir u baie maklik sal maak om die teorie die res van die kursus baie maklik wees om te verstaan Prostaatanker : Prostaatanker kan baie maklik opgespoor word deurdat die prostaatklieer Die meganisme van oorkruising kan baie maklik met kleimodelle voorgestel word meting . Dit is weereens baie maklik by die natuurwetenskappe bereikbaar ...	Taalkommissie-korpus.Nie-fiksie.Akademies.Studiegids.VerskeieWetenskappe3	28

Onder die opskrif "Trefslae in groepe" kan jy die resultate groepeer volgens een bepaalde kategorie, byvoorbeeld volgens outeursnaam, woord links, woordsoort van die trefslag, ens. Hieronder verskyn die resultate gesorteer volgens "woord links". Daaruit blyk dat "is" die hoogste voorkoms voor die string "baie maklik" in die korpora het, byvoorbeeld in die sin Dit is baie maklik om te sê wat die meervoud van hond is.

Soek Korpusinligting

Eenvoudig Uitgebreid Gevorderd Kuldig || Resultate Dokument

Soekopdragte:

#	Soekopdrag	binne	Metadatafilters	Groepering	Status	Trefslae	Dokumente	
1	[[word="baie"]][[word="maklik"]]	document		wordleft:word	KLAAR	359	66	VERANDER

Soekresultate:

Trefslae Dokumente Trefslae in groepe Dokumente in groepe

woord links

Toon 50 per bladsy << < 1 2 3 > >> Bladsy 1 van 3 Gaar

Stoor resultate

Groep ▼	Trefslae ▲
is	46
dit	41
kan	33
Dis	15
wat	12
sal	7
hulle	6

Onder die opskrif "Dokumente in groepe" kan jy die brontekste groepeer volgens drie groot kategorieë: outeur, publikasiedatum en korpus ("versamelingnaam"). Hieronder is die resultate gesorteer volgens korpus. Daaruit blyk dat die meeste trefslae aangetref word in 19 dokumente van die PUK/Protea Boekhuis-korpus.

Soek Korpusinligting

Eenvoudig Uitgebreid Gevorderd Kuldig || Resultate Dokument

Soekopdragte:

#	Soekopdrag	binne	Metadatafilters	Groepering	Status	Trefslae	Dokumente	
1	[(word="baie")]((word="maklik"))	document		field.CollectionName	KLAAR	359	66	VERANDER

Soekresultate:

Trefslae | Dokumente | Trefslae in groepe | Dokumente in groepe

Versamelingnaam

Toon 50 per bladsy << < 1 > >> Bladsy 1 van 1 Gaar

Stoor resultate

Groep ▲ ▼	Dokumente ▲ ▼
PUK/Protea Boekhuis-korpus	19
NWU/LAPA-korpus	13
RSG-nuuskorpus	7
WatKykJy-Korpus	4
NCHLT- Afrikaanse Korpus	1
Taalkommissie-korpus.Nie-fiksie.Nie-akademies.Boeke Ander	1
Taalkommissie-korpus.Nie-fiksie.Nie-akademies.Boeke Algemeen	1

Onder beide die opskrifte “Trefslae in groepe” en “Dokumente in groepe” kan jy op die groen balk van ’n soekresultaat klik. Daar verskyn dan nog ’n knoppie: “vertoon gedetailleerde dokumente in hierdie groep”.

Soek Korpusinligting

Eenvoudig Uitgebreid Gevorderd Kuldig || Resultate

Soekopdragte:

#	Soekopdrag	binne	Metadatafilters	Groepering	Status	Trefslae	Dokumente	
1	[(word="baie")]((word="maklik"))	document		wordleft:word	KLAAR	359	66	VERANDER

Soekresultate:

Trefslae | Dokumente | Trefslae in groepe | Dokumente in groepe

woord links

Toon 50 per bladsy << < 1 2 3 > >> Bladsy 1 van 3 Gaar

Voer uit

Groep ▲ ▼	Trefslae ▲ ▼
is	46
dit	41
kan	33
Dis	15
wat	12
sal	7

Vertoon gedetailleerde dokumente in hierdie groep

As jy op hierdie knoppie klik, word ’n nuwe soekopdrag uitgevoer, wat slegs die soekresultate toon van die item waarop geklik word. Hieronder word die soekopdrag en die eerste resultate getoon indien jy soek na die trefslae waarin wat links staan van baie maklik, soos in lang gras wat baie maklik aan die brand slaan.

Soek Korpusinligting

Eenvoudig Uitgebreid Gevorderd Kundig || Resultate

Soekopdragte:

#	Soekopdrag	binne	Metadatafilters	Groepering	Status	Trefslae	Dokumente	
1	[(word="baie")][(word="maklik")]	document		wordleft:word	KLAAR	359	66	VERANDER X
3	[word="(?!-)wat"][(word="baie")][(word="maklik")]			-	KLAAR	12	10	VERANDER X

Soekresultate:

Trefslae | Dokumente | Trefslae in groepe | Dokumente in groepe

Toon 50 per bladsy

Bladsy 1 van 1 Gaar

Voer uit

Konteks links	Trefstuk ▲ ▼	Konteks regs	lemma ▲ ▼	woordsoort ▲ ▼
... baie droog met lang gras	wat baie maklik	aan die brand kan slaan ...	wat baie maklik	VNW.betreklik TW.hoofelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... Dit is 'n program	wat baie maklik	is om aan te leer ...	wat baie maklik	VNW.betreklik TW.hoofelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... van beserings by die sportman	wat baie maklik	voorkom ... BEGINSLS VAN BEHANDELING ...	wat baie maklik	VNW.betreklik TW.hoofelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... is 'n hoogs aansteeklike ooginfeksie	wat baie maklik	versprei . Dit kan veroorsaak ...	wat baie maklik	VNW.betreklik TW.hoofelwoord.bywoord.onbepaald B.NW.stellend.predikatief
... Dis die soort ding	wat baie maklik	deur ander fotograawe nageaap sou ...	wat baie maklik	VNW.betreklik TW.hoofelwoord.bywoord.onbepaald B.NW.stellend.predikatief

4.2 Uitgebrei

4.2.1 Soekopdrag

Die uitgebreide soekmetode stel 'n mens in staat om metadata te filter en volgens woorde, lemmas en/of woordsoorte te soek. Om die metadata te filter, kan jy 'n aantal "reëls" instel. So kan jy byvoorbeeld in een bepaalde korpus soek, of jy kan 'n publikasiedatum of outeursnaam kies. Jy kan veelvuldige filters stel deur regs op die +-teken te klik. Die onderstaande soekopdrag is na mooi in tekste wat in 2008 gepubliseer is.

Soek Korpusinligting

Eenvoudig **Uitgebreid** Gevorderd Kundig || Resultate Dokument

Metadatafilters

Pas hierdie reëls toe:

Publikasiedatum is 2008

Groepeer trefslae per

Soek binne dokument

woord Kassensitief

lemma Kassensitief

woordsoort

Skei bondelnavrae

Soek Herstel

Normaalweg is soekopdragte nie kassensitief nie. As jy 'n kassensitiewe soekopdrag wil skep, moet jy die blokkie "kassensitief" merk.

4.2.2 Resultate

Die resultate word in die “resultate”-venster vertoon. Soos in afdeling 4.1.2 aangedui, kan jy die resultate sorteer en aflaai.

The screenshot shows the 'Korpusinligting' search interface. At the top, there's a navigation bar with 'Soek' and 'Korpusinligting'. Below it, a menu includes 'Eenvoudig', 'Uitgebreid', 'Gevorderd', 'Kundig', and 'Resultate' (which is selected). The main content area is divided into two sections: 'Soekopdragte:' and 'Soekresultate:'.

Soekopdragte: A table with columns: #, Soekopdrag, binne, Metadatatfilters, Groepering, Status, Trefslae, Dokumente, and a 'VERANDER' button. Two queries are listed:

#	Soekopdrag	binne	Metadatatfilters	Groepering	Status	Trefslae	Dokumente	VERANDER
1	[[word="baie"]][[word="maklik"]]	document		field:CollectionName	KLAAR	359	66	VERANDER
4	[lemma="(?)mooi"]	document	PublicationDate="2008"	-	KLAAR	719	18	VERANDER

Soekresultate: This section includes filters for 'Trefslae', 'Dokumente', 'Trefslae in groepe', and 'Dokumente in groepe'. It shows 'Toon 50 per bladsy' and a pagination control with buttons for '<<', '<', '1', '2', '3', '4', '5', '6', '...', '>', '>>', and a 'Toggle titels' button. A 'Bladsy 1 van 15' indicator and a 'Gaan' button are also present. A 'Stoor resultate' button is located below the pagination.

The main results table has columns: 'Konteks links', 'Trefstuk', 'Konteks regs', 'lemma', and 'woordsoort'. It displays a list of search results for the query '[lemma="(?)mooi"]:


Konteks links	Trefstuk	Konteks regs	lemma	woordsoort
... een van die wêreld se	mooiste	skepe - die Esmeralda in ...	mooi	B.NW.stellend.attritief
... Die wêreld word al	mooier	, ná die eindelose grasvlaktes. ...	mooi	WW.
... het . Dit was die	mooiste	plek . Wild het daar ...	mooi	B.NW.stellend.attritief
... vasgesit. Altesaam dertien bulle met	mooi	tande is geskiet . Die ...	mooi	
... die houtploeg aangelê , 'n	mooi	oes ingesamel en verder getrek ...	mooi	S.NW.soortnaam.enkelvoud.nominatief.basis
... frisser en hy het 'n	mooier	teerheid by Willem Bessemer teenoor ...	mooi	B.NW.stellend.attritief
... tot stilstand bring om eers	mooi	te kyk . Na die ...	mooi	WW.
... eers loop oë wys en	mooi	kyk . " Toe jaag ...	mooi	
... kry . En voor Naas	mooi	weet , het sy haar ...	mooi	S.NW.soortnaam.meenvoud.nominatief.basis
... vlam. Dit is nog nie	mooi	lig nie , toe hy ...	mooi	
... tref hom weer hoe volmaak	mooi	sy is . Sy laat ...	mooi	WW.
... hulle joue , Naas .	Mooi	loop en lig loop vir ...	Mooi	S.NW.soortnaam.enkelvoud.nominatief.basis

In die soekopdrag is die metadatatfilter geskep: Slegs dokumente wat in 2008 geskep is, word deursoek. Let op dat die lemma mooi gesoek is. Daarom is die vormlike variante soos mooie, mooier en mooiste ook gevind.

Bokant die nuwe soekopdrag is die vorige soekopdrag nog sigbaar. As jy daarop klik, kan jy die resultate van die soekopdrag opnuut bekijk.

4.2.3 Bondelnavrae

In die uitgebreide soekmodus is dit moontlik om “bondelnavrae” (batch queries) uit te voer. Dit wil sê dis moontlik om verskillende soekopdragte gelyktydig uit te voer. As jy 'n lang lys woorde, lemmas of woordsoorte het wat jy in Korpusportaal wil naslaan, hoef jy nie elke opdrag apart in te voer nie.

Deur op die -knoppie regs van woord, lemma of woordsoort te klik, kan jy 'n stuk teks oplaai (plat teks, NIE 'n Word-dokument nie) met 'n lys soekopdragte. Die items in die lys moet dieselfde soort wees (woorde OF lemmas OF woordsoorte).

The screenshot shows the VIVA search interface. At the top, there is a blue header with a question mark icon, the text 'Soek', and 'Korpusinligting'. Below the header, there are navigation tabs: 'Eenvoudig', 'Uitgebreid' (highlighted in red), 'Gevorderd', and 'Kundig'. The VIVA logo and 'VIRTUELE INSTITUUT VIR AFRIKAANS' are in the top right. A section titled 'Metadatafilters' contains three filter fields: 'woord' with the value 'woord', 'lemma' with the value 'lemma', and 'woordsoort'. Each field has a 'Kassensitief' checkbox. A 'Skei bondelnavrae' checkbox is located below the filters. At the bottom are 'Soek' and 'Herstel' buttons. A black circle highlights the empty input area of the 'woord' field.

In die onderstaande voorbeeld word lemmas gelaai.

This screenshot shows the same VIVA search interface as above, but with the 'lemma' filter field populated with suggestions: 'leeu' and 'renoster'. The 'Skei bondelnavrae' checkbox is now checked. The 'Soek' and 'Herstel' buttons remain at the bottom.

Die blokkie “skei bondelnavrae” word altyd gemerk. Dit wil sê elke reël in die soekvenster word as 'n aparte navraag hanteer. Die resultatebladsy lyk so:

Soek Korpusinligting

Eenvoudig Uitgebreid Gevorderd Kundig || Resultate

VIVA
VIRTUELE INSTITUUT VIR AFRIKAANS

Soekopdragte:

#	Soekopdrag	binne	Metadatalfilters	Groepering	Status	Trefslae	Dokumente	
1	[lemma="(?)olifant"]	document		-	KLAAR	2871	111	VERANDER X
2	[lemma="(?)leeu"]	document		-	KLAAR	6535	162	VERANDER X
3	[lemma="(?)renoster"]	document		-	KLAAR	1412	73	VERANDER X
4	[lemma="(?)buffel"]	document		-	KLAAR	837	108	VERANDER X
5	[lemma="(?)luiperd"]	document		-	KLAAR	1158	89	VERANDER X

Soekresultate:

Trefslae: **Dokumente** Trefslae in groepe Dokumente in groepe

Toon 50 per bladsy << < 1 2 3 4 5 6 ... > >> Toggle titels Bladsy 1 van 17 Gaan

Stoor resultate

Konteks links	Trefstuk ▲ ▼	Konteks regs	lemma ▲ ▼	woordsoort ▲ ▼
... Werker van Limpopo is deur ... besig toe hy deur die ... bestel. Twee Amerikaanse maatskappye vervaardig ... veroorsaak dat siektdraende diere soos ... van koedoes, nyala's,	buffels buffels Buffel buffels buffels	op Hoedspruit doodgetrap . Die ... aangeval is . Krieket : ... en Cougar gepantserde voertuie met ... op plase beland . Die ... en kameelperde wat in die ...	buffel buffel Buffel buffel buffel	S.NW.soortnaam.meervoud.nominatief.basis S.NW.soortnaam.enkelvoud.nominatief.basis WW. S.NW.soortnaam.meervoud.nominatief.basis S.NW.soortnaam.meervoud.nominatief.basis

As jy die blokkie “skei bondelnavrae” se merkie verwyder, word al die items in die bondelnavraag na 'n enkele soekopdrag omgeskakel.

Soek Korpusinligting

Eenvoudig Uitgebreid Gevorderd Kundig || Resultate

VIVA
VIRTUELE INSTITUUT VIR AFRIKAANS

Soekopdragte:

#	Soekopdrag	binne	Metadatalfilters	Groepering	Status	Trefslae	Dokumente	
6	[[lemma="(?)olifant" lemma="(?)leeu" lemma="(?)renoster" lemma="(?)buffel" lemma="(?)luiperd"]]	document		-	KLAAR	12813	217	VERANDER X

Soekresultate:

Trefslae: **Dokumente** Trefslae in groepe Dokumente in groepe

Toon 50 per bladsy << < 1 2 3 4 5 6 ... > >> Toggle titels Bladsy 1 van 257 Gaan

Stoor resultate

Konteks links	Trefstuk ▲ ▼	Konteks regs	lemma ▲ ▼	woordsoort ▲ ▼
... , hiënas , wildehonde , ... , wildehonde , leeus en ... , selfs op die rûe van ... Lotz-dieretuin in Pole gevrek. Die ... sê die heining is deur	leeus luiperds olifante renoster olifante	en luiperds word deur die ... word deur die regulasies gedek. ... in die huwelik getree wat in 1959 in die ... plaatgetrap omdat dit nie geëlektrifiseer ...	leeu luiperd olifant renoster olifant	S.NW.soortnaam.meervoud.nominatief.basis S.NW.soortnaam.meervoud.nominatief.basis S.NW.soortnaam.meervoud.nominatief.basis S.NW.soortnaam.enkelvoud.nominatief.basis B.NW.stellend.attributief

Die resultate kan op dieselfde manier geraadpleeg word soos beskryf in afdeling 4.2.1.

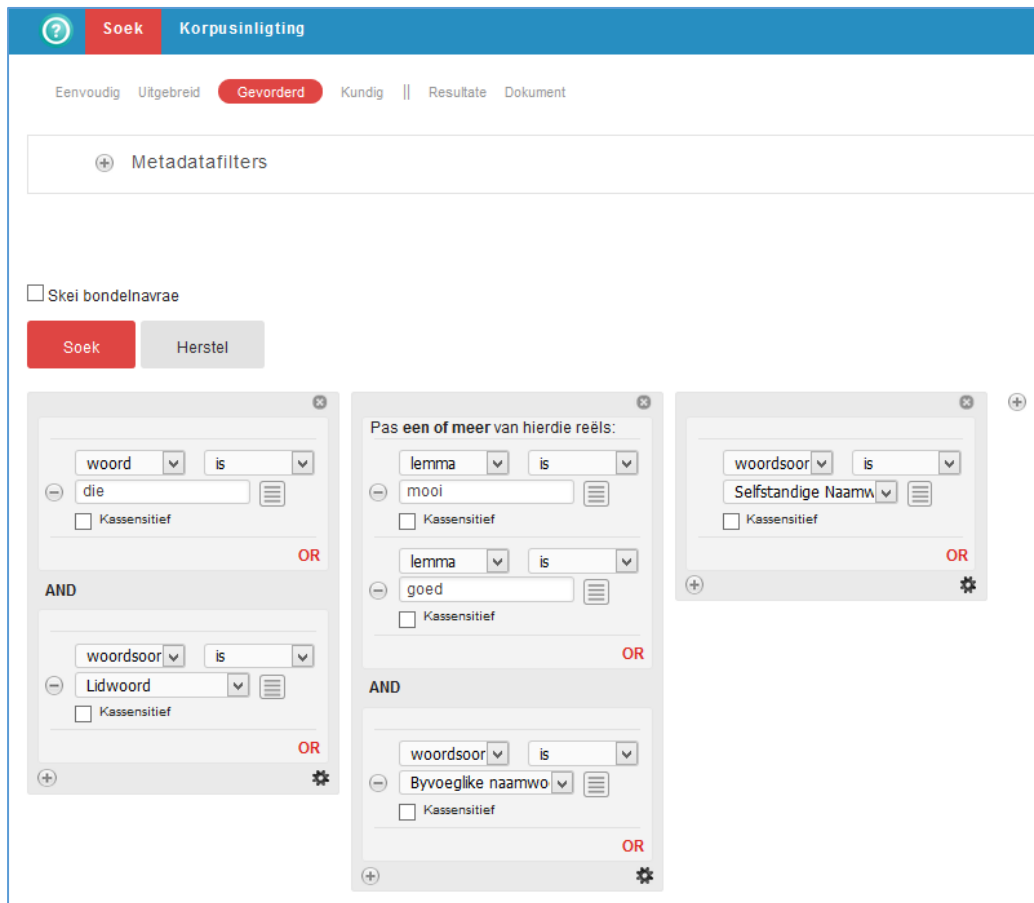
4.3 Gevorderd

4.3.1 Soekopdrag

Die gevorderde soekmetode stel jou in staat om metadata te filter, soos uiteengesit in afdeling 4.2.1. Hierdie invoermetode is 'n kombinasie van die vorige twee. Soos in die eenvoudige soekmodus is dit moontlik om een of meer soekterme te definieer. Soos in die uitgebreide soekmodus is dit moontlik om vir elke item aan te dui of jy in die presiese woordvorm, die lemma of die woordsoort belangstel. Daar is in hierdie soekmodus ook kombinasies moontlik. Elke blok dui een element van die soekopdrag aan. In daardie blok kan jy inligting oor die woordvorm, die lemma en/of die woordsoort gee. Deur op die +-simbool regs te klik, kan jy veelvuldige elemente aan die soekopdrag koppel. Die onderstaande soekopdrag soek na kombinasies van die as lidwoord, gevolg deur mooi of goed as adjektief, gevolg deur 'n selfstandige naamwoord.

Deur op die +-simbool links onder 'n blokkie te klik, kan jy 'n AND-beperking oplê. So kan jy aandui dat jy soek na die word die EN dat dit 'n lidwoord moet wees. Deur op OR te klik, kan jy 'n OR-beperking oplê. So kan jy byvoorbeeld aandui dat jy op soek is na die lemma mooi OF die lemma goed.

Via die ⚙️-knopie kan jy bykomende voorwaardes stel, byvoorbeeld om slegs trefslae aan die begin van 'n sin weer te gee.



4.3.2 Resultate

Die resultate word in die "resultate"-venster vertoon. Soos in afdeling 4.1.2 aangedui, kan jy die resultate sorteer en aflaai. Die onderstaande figuur toon die gegroepeerde trefslae, waaruit dit blyk dat die string die goeie nuus die trefslag met die hoogste voorkoms het wat aan die soekopdrag hierbo voldoen.

Soek Korpusinligting

Eenvoudig Uitgebreid Gevorderd Kuldig || **Resultate** Dokument

Soekopdragte:

#	Soekopdrag	binne	Metadatalfilters	Groepering	Status	Trefslae	Dokumente	
1	[[word="baie"]][[word="maklik"]]	document			KLAAR	359	66	VERANDER X
4	[lemma="(?)mooi"]	document	PublicationDate="2008"	-	KLAAR	719	18	VERANDER X
10	[word="(?)die" & pos="(?)LD."][[lemma="(?)mooi" lemma="(?)goed" & pos="(?)B.NW."][[pos="(?)S.NW."]]	document		hit word	KLAAR	1121	210	VERANDER X

Soekresultate:

Trefslae | Dokumente | Trefslae in groepe | **Dokumente in groepe**

hit woord

Toon 50 per bladsy


<< < 1 2 3 4 5 6 ... > >>

Bladsy 1 van 12 Gaan

Stoor resultate

Groep ▲ ▼	Trefslae ▲ ▼
die goeie nuus	75
Die goeie nuus	50
die mooiste meisie	23
die goeie lewe	22
die Goeie Hoop	21
die goeie dinge	18
die mooiste vrou	17
die goeie tye	16

4.3.3 Bondelnavrae

Dit is ook in hierdie soekmodus moontlik om bondelnavrae uit te voer deur 'n lys met soekopdragte via die -knoppie te laai. Hierdie funksie werk op 'n soortgelyke manier as beskryf in afdeling 4.2.3.

4.4 Kuldig

4.4.1 Soekopdrag

"Kuldig" is die gevorderdste soekmodus. Die metadata kan gefilter word soos uiteengesit in afdeling 4.2.1. Daarbenewens bestaan die koppelvlak uit 'n teksveld waarin jy 'n soekopdrag in **Corpus Query Language (CQL)** kan gee, CQL maak dit moontlik om komplekse soekopdragte te definieer. Hiervoor het jy wel kennis nodig van die CQL-formaat en die gebruikte etikette vir woordsoortannotasie. Die voordeel is dat jy meer beheer het oor die patrone wat jy opsoek. As jy op die vraagtekenikon klik, kry jy 'n aantal voorbeelde van soekopdragte in CQL-formaat.

5 Corpus Query Language (CQL)

Corpus Query Language (CQL) is 'n formele soektaal waarmee 'n mens in geannoteerde korpora kan soek. Afgesien van Korpusportaal maak ander soekenjins ook daarvan gebruik. Soos in afdeling 4 uiteengesit is, maak CQL dit moontlik om meer komplekse soekopdragte te formuleer. Hiervoor het jy wel bietjie kennis nodig van die CQL-formaat en die gebruikte etikette vir woordsoortannotasie. Die voordeel is dat jy meer beheer het oor die patrone wat jy opsoek. Hierdie afdeling gee 'n inleiding tot die CQL-formaat, soos van toepassing op die Korpusportaal.² 'n Oorsig van alle afkortings en operatore is te vinde in die Addendum.

Jy kan CQL as invoermetode in die **soekmodus "kundig"** van Korpusportaal (kyk afdeling 4) gebruik. Indien jy 'n ander soekmodus gebruik, word die vraag outomaties in 'n CQL-navraag omskep. Dit kan jy in die resultatevenster (onder "soekopdragte") sien. So kan jy vertrou raak met die formulering sonder dat jy aktief daarmee hoef om te gaan. Jy kan ook outomaties gegenereerde CQL-navrae na die soekmodus "kundig" kopieer en verder verfyn. Dit is soms makliker as om die navrae van nuuts af saam te stel. Dit is tog goed om 'n aantal basiese beginsels van CQL onder die knie te hê voor jy hiermee begin. Dit word in die volgende afdeling bespreek.

Waarskuwing: Dit is belangrik dat jy die **regte aanhalingstekens** in CQL gebruik, anders gaan jy 'n foutboodskap kry. As jy soekopdragte uit woordverwerkingsprogramme soos MS Word of OpenOffice kopieer, moet jy hiermee rekening hou. Soekopdragte wat uit plat teks gekopieer word, behoort geen probleem te skep nie.

5.1 Eenvoudige soekopdragte

In hierdie afdeling word 'n paar eenvoudige CQL-soekopdragte toegelig. Soortgelyke soekopdragte word die boustene van meer komplekse soekinstryksies.

5.1.1 Enkele basiese beginsels

- Elke element (word, lemma en/of woordsoort per item) moet tussen **vierkantige hakies** gedefinieer word: []
- Jy kan die tipe element wat jy soek definieer deur **word=**, **lemma=**, of **pos=** daarvoor te tik.³
- As jy 'n woord, lemma of woordsoortetiket wil definieer, moet jy dit tussen **reguit (enkele ' ' of dubbele " ") aanhalingstekens** plaas.

5.1.2 Soek na 'n woord, lemma of woordsoort

- Soek alle instansies van die word lekker. **[word="lekker"]** of **["lekker"]** of **"lekker"**
Hierdie soekopdrag soek na die presiese woordvorm lekker, maar is nie kassensitief nie. Sinne met Lekker word dus ook in die resultate vertoon, maar sinne met lekkerder of lekkerste nie.
As jy slegs na die woordvorm soek, hoef jy nie die etiket "word=" en die hakies te spesifiseer nie.
- Soek alle instansies van die lemma lekker. **[lemma="lekker"]**

² As jy na dese meer te wete wil kom van CQL, kan jy aanlyn na CQL-handleidings soek. Kyk byvoorbeeld <https://www.sketchengine.co.uk/corpus-querying>.

³ POS staan vir die Engelse "part of speech" (woordsoort, woordklas).

Hierdie soekinstruksie soek na sinne waarin lekker en afleidings daarvan voorkom, soos lekker, lekkerder, lekkerste, ens.

- Soek konstruksies met 'n bepaalde lidwoord. **[pos="LID.bepaald"]**
Hierdie soekopdrag soek na items wat die etiket "LID.bepaald" het, d.w.s. die en Die.
- Met die &-simbool kan jy meer eienskappe van een element definieer, soos die lemma en die woordsoort. So kan jy na groot as attributiewe adjektief soek. Op dié manier kan jy alle gevalle van groot as predikatiewe adjektief uitsluit.
[lemma="groot" & pos="B.NW.stellend.attributief"]
Die soekresultate bevat trefslae soos 'n groot deurbraak, groot kinders, die groot familie.

5.1.3 Soek na naasliggende woorde, lemmas en/of woordsoorte

Jy kan na verskillende elemente, soos twee opeenvolgende woorde of kombinasies van woorde, lemmas of woordsoorte soek deur die verskillende elemente tussen vierkantige hakies te definieer.

- Soek na gevalle van baie lekker. **[word="baie"] [word="lekker"]** of **["baie"] ["lekker"]** of **"baie" "lekker"**
Die resultate bevat konstruksies met baie lekker, Baie lekker en BAIE LEKKER. Ook in hierdie geval is die hakies en die etiket "word=" opsioneel.
- Jy kan ook verskillende soorte elemente kombineer. Jy kan so soek na konstruksies wat bestaan uit die woord die, 'n attributiewe adjektief en die lemma man.
[word="die"] [pos="B.NW.stellend.attributief"] [lemma="man"]
Voorbeelde van soekresultate is die 40-jarige man, die gesogte man, Die gewapende mans.
- Jy kan veelvuldige eienskappe per element definieer. So kan jy soek na gevalle van geveg as selfstandige naamwoord waar dit gevolg word deur het.
["geveg" & pos="S.NW.soortnaam.enkelvoud.nominatief.basis"] ["het"]
Die resultaat bevat konstruksies soos die geveg het begin; konstruksies met geveg as werkwoord (bv. teen wie hy geveg het) is hierby uitgesluit.

5.1.4 Soek na nienaasliggende konstruksies

Jy kan na konstruksies soek waarin die betrokke elemente nie naas mekaar voorkom nie, soos die kombinasie van 'n werkwoord en 'n partikel. Met behulp van leë vierkantige hakies [] kan jy aandui hoeveel items daar tussen die elemente moet voorkom.

- Soek na konstruksies met die man waarin daar 'n willekeurige item tussen die en man voorkom.
"die" [] "man"
Die resultaat bevat konstruksies soos die ander man, die 50-jarige man, die jong man.

Jy kan aandui dat daar meer elemente tussen die bepaalde items mag staan deur meer hakies te gebruik, bv. **"die" [] [] "man"** of **"die" [][][] "man"**. Vir soekopdragte waarin die elemente ver uit mekaar kan staan, is dit egter nie 'n effektiewe werkwyse nie. 'n Ander manier is om die afstand tussen die elemente met krulhakies {} aan te dui. So kan die soekopdrag **"die" [][][] "man"** ook geformuleer word as **"die" []{3} "man"**.

- Soek na konstruksies met aflaai, waarin twee items tussen die werkwoord en die partikel staan.
"laai" []{2} [lemma="af" & pos="U.partikel.wv."]

Die resultaat bevat konstruksies soos toe gaan laai ek haar af, laat dit nou af, Joe laai sy dogter af.

Jy kan ook 'n interval definieer, byvoorbeeld dat daar tussen twee en vyf elemente tussen die werkwoord en die partikel staan. "laai" `[{2,5} [lemma="af" & pos="U.partikel.ww."]` Dit gee trefslae soos die ander man, laai my asseblief by die kantoor af, laai jou weer veilig af.

5.2 Boolese operatore

Boolese operatore is **logiese operatore** wat gereeld in die informatika en wiskunde gebruik word. CQL maak ook daarvan gebruik. Die drie operatore in CQL is AND ("en"), OR ("of") en NOT ("nie"). Dit stel jou in staat om onderskeidelik konjunksie, disjunksie en negasie uit te druk. Hiermee kan jy meer gevorderde soekopdragte formuleer.

5.2.1 AND

Die gebruik van die AND-operator is in afdeling 5.1.2 bespreek. Met die **&**-simbool kan jy 'n AND-verhouding definieer. So kan jy byvoorbeeld soek na kombinasies van lemmas en woordsoortannotasies. So kan jy aan die hand van die soekopdrag `[lemma="groot" & pos="B.NW.stellend.predikatief"]` die gevalle van groot as predikatiewe adjektief in Korpusportaal opspoor. Voorbeelde van trefslae is die impak sal groot wees, die risiko is te groot.

5.2.2 OR

Met die OR-operator kan jy in een soekopdrag na meer as een element op 'n slag soek. Die OR-verhouding word uitgedruk met die **|**-simbool. So soek die CQL-soektog `["klein"|"groot"]` na gevalle van die woord klein of die woord groot. Voorbeelde van trefslae is 'n groot gedeelte, klein entrepreneurs. 'n Ander formulering wat dieselfde resultaat sal oplewer, is `[word="klein|groot"], [word="klein" | word="groot"]`.

Die OR-funksie werk nie net met woorde nie; jy kan ook tegelyk na verskillende lemmas of woordsoorte soek. Die soekopdrag `[pos="VNW.aanwysend|VNW.vraend"]` lewer resultate op wat 'n aanwysende (bv. hierdie, dié) of vraende voornaamwoord (bv. wie, hoe, wat) bevat.

Vir hierdie soekopdrag is daar ook 'n ander skryfwyse moontlik: `[pos="VNW.aanwysend"|pos="VNW.vraend"]`. Die notasie `[pos="VNW.aanwysend"|"VNW.vraend"]` is nie 'n alternatief nie, want die soektog is dan na VNW.aanwysend as woordsoort, maar na VNW.vraend as woord.

5.2.3 NOT

Die NOT-funksie stel jou in staat om bepaalde elemente by 'n soekopdrag uit te sluit. Dit kan jy in CQL doen deur 'n **uitroep-teken voor die is-gelyk-aan-teken (!=)** te plaas. Die soektog `["suid" [word != "Afrika"]` gaan soek na trefslae waarin die woord suid gevolg word deur 'n ander woord, maar die tweede woord mag nie Afrika wees nie. Die resultate bevat byvoorbeeld suid van die Sahara.

Dit is moontlik om aan te dui dat een element aan voorwaarde A moet voldoen, maar nie aan voorwaarde B nie. Die soekopdrag `["suid" [word != "Afrika" & pos="EIE.eienaam.enkelvoud.nominatief.basis"]` sal trefslae lewer waar die woord suid wat nie deur Afrika gevolg word nie, maar wel deur 'n eienaam. Die resultate bevat onder andere die skryffoute Suid Korea, Suid Soedan, maar geen kombinasies soos suid van ... nie

5.3 Vaste uitdrukings

Vaste uitdrukings (Eng. regular expressions, regex) stel jou in staat om 'n soekpatroon te definieer. Net soos die Boolese operatore is vaste uitdrukings nie bepaald vir CQL geskep nie, maar daar bestaan verskillende toepassing wat daarvan gebruik maak. 'n Eenvoudige toepassing daarvan is die "vind en vervang"-funksie in teksverwerkers. Daarmee kan jy soek na spesifieke patrone (of strings) in 'n teks en kan jy jou oorheersstring (wild card) gebruik om na verskillende patrone tegelyk te soek. In Korpusportaal kan jy vaste uitdrukings gebruik om na woorde te soek, maar ook om na lemmas of woordsoorte te soek.

Die gewone letters en syfers in 'n vaste uitdrukking kom ooreen met dieselfde tekens in die teks. So sal die soektog **[word="3-duisend"]** die woord 3-duisend in die korpus teks opspoor.⁴

In hierdie afdeling word die belangrikste **spesiale tekens** toegelig wat jy in kombinasie met CQL kan gebruik.

5.3.1 Oorheersstring vir 'n willekeurige karakter

'n **Punt** (.) staan vir 'n willekeurige teken. Op dié manier lewer die soekopdrag **["b.l"]** onder andere bal, bol, bil, BOL op.

Vaste uitdrukings is by verstek **nie kassensitief nie**.

5.3.2 Groepering van karakters en patrone

Met behulp van **vierkantige hakies ([])** kan jy 'n karakterklas, d.w.s. 'n lys van moontlike tekens, definieer. Die vaste uitdrukking **["b[aie]"]** soek na bal, bil en bel (maar nie na bol of bul nie).

In afdeling 5.2.2 word die OR-operator | beskryf. Jy kan dit gebruik nie net om **opsionaliteit** tussen volledige woorde of woordsoorte aan te dui nie. As jy byvoorbeeld op soek is na alle gevalle van meervoudige besitlike voornaamwoorde, kan jy die volgende soekopdrag gebruik: **[pos="VNW.eerste.meervoud.besitlik"|pos="VNW.tweede.meervoud.besitlik"|pos="VNW.derde.meervoud.besitlik"]**. 'n Kompakter alternatief om die disjunksie weer te gee, is **[pos="VNW.(eerste|tweede|derde).meervoud.besitlik"]**, waarin jy die verskille tussen die etikette tussen ronde hakies aandui.

'n Ander voorbeeld is **["ou(m|p)a"]**, wat ouma sowel as oupa oplewer en dus ooreenkom met ["ouma|oupa"].

5.3.3 Kwantifikasie

Die **kwantore +, ? en *** dui aan hoe gereeld 'n bepaalde teken in die soekpatroon sou voorkom.

- Die **vraagteken (?)** Dui aan dat die voorafgaande karakter 0 of 1 keer kan voorkom. **["voorbeeld?"]** soek na gevalle van voorbeeld en voorbeelde. Die kwantore kan soos die OR-operator in kombinasie met groeperings in ronde hakies gebruik word. **["seun(tjie)?"]** lewer seun en seuntjie op.
- Die **plusteken (+)** dui aan dat die voorafgaande karakter (of groep karakters) 1 of meer keer kan voorkom. **["do+r"]** soek na dor, door, door, ens.

⁴ In ander toepassings hou vaste uitdrukings deurgaans nie rekening met woordgrense nie. Daarom sal die vaste uitdrukking **"bal"** in sommige toepassing ook *ballet* oproep. In die kombinasie met CQL word daar egter rekening gebou met woordgrense: Die vierkantige hakie dui immers 'n woordgrens aan.

- Die **asterisk (*)** dui aan dat die voorafgaande karakter (of groep karakters) 0, 1 of meer keer kan voorkom. `["do*r"]` soek na dr, dor, door, dooor, ens.

'n Gereelde vaste uitdrukking is `[".*"]` wat alle teks opspoor. Die soektog is ook nuttig om algemene woordsoortannotasiesoektogte te doen. As jy byvoorbeeld op soek is na alle voornaamwoorde, kan jy die soekopdrag `[pos="VNW.*"]` gebruik.

Jy kan die kwantore nie net binne 'n element gebruik nie, maar ook om 0, 1 of veelvuldige voorkomste van dieselfde elemente te definieer. So kan jy soek na konstruksies met een of meer werkwoorde met behulp van die soekopdrag `[pos="WW.*"]+`.

Jy kan ook gebruik maak van **krulhakies** om uit te druk hoe gereeld 'n bepaalde element kan voorkom. So soek die vaste uitdrukking `["do{2}r"]` na door, `["do{1,2}r"]` soek na dor en door. `["hallo{2,}"]` soek na konstruksies met twee of meer o's, soos halloo en hallooo.

Hierdie manier van kwantifiseer kan jy in CQL ook vir die volledige element gebruik.

`[pos="WW.*"]{2}` soek na konstruksies met twee werkwoorde, soos opgelos word. `[pos="WW.*"]{2,3}` soek na konstruksies met twee of drie werkwoorde, soos opgelos word en probeer insmokkel het. `[pos="WW.*"]{2}` soek na konstruksies met twee of meer werkwoorde.

5.3.4 Ontsnapteken

Soms wil jy soek na die **werklike weergawe van die spesiale tekens** in die korpus. Dit kan jy doen deur 'n agtertoeskuinsstreep (\) voor die betrokke teken te plaas. `["\""]` soek na 'n punt, `["?\""]` na 'n vraagteken, `["\""]` na 'n hakie. Die agtertoeskuinsstreep self kan jy opspoor met `["\\"]`.

5.3.5 Kassensitiewe soekopdragte

In die vorige afdelings is daar aangedui dat CQL-soektogte nie kassensitief is nie. Indien jy wel 'n kassensitiewe opdrag wil formuleer, kan jy die operator (?-i) tot die kassensitiewe deel van die soektog toevoeg.⁵

`["(?-i)Lekker"]` soek na die voorkoms van Lekker. Sinne met lekker of LEKKER word nie in dié resultate vertoon nie.

In soekopdragte met veelvuldige elemente, of indien die OR-operator gebruik word, het die (?-i)-toevoeging nie betrekking op die volledige soekopdrag nie. `["(?-i)Lekker"]"dag"]` gee konstruksies met Lekker, dag, Dag en DAG as resultaat. As jy die tweede element van die disjunksie ook kassensitief wil maak, moet jy dit soos volg doen: `["(?-i)Lekker"]"(?-i)dag"]` of `["(?-i)(Lekker|dag)"]`. Die soektog soek slegs na Lekker en dag.

⁵ Die tegenhanger van die hooflettergevoelige operator (?-i) is (?i). Aangesien navrae sonder (?-i) by verstek nie kassensitief is nie, is dit oorbodig om (?i) toe te voeg. Die soekopdrag `["Lekker"]` lewer dieselfde resultaat as die opdrag `["(?i)Lekker"]`.

6 Addendum

6.1 Korpora in Korpusportaal

Daar is drie vlakke van toegang tot VivA se korpusportaal:

- **OOP:** Die korpora in hierdie groep is vrylik aan alle VivA-gebruikers vir navorsingsdoeleindes beskikbaar.
- **OMVATTEND:** Die korpora in hierdie groep sluit ook die korpora in die OOP-vlak in. Toegang is beperk tot bona fide-navorsers. Verkry toegang tot hierdie korpora deur die [aanlyn aansoekvorm](#) te voltooi.
- **EKSKLUSIEF:** Die korpora in hierdie groep is slegs beskikbaar aan bona fide-navorsers wat kan aantoon dat die korpora vir 'n spesifieke projek benodig word. Verkry toegang tot hierdie korpora deur die [aanlyn aansoekvorm](#) te voltooi.

Vlak	Naam	Aantal woorde	Aantal tekseenhede
OOP	NCHLT- Afrikaanse korpus 1.0	2 229 214	2 489 716
OMVATTEND	Afrikaanse Leipzig-korpus 1.0	28 776 800	32 269 153
OMVATTEND	NWU/Maroela Media-korpus 1.2	9 173 430	10 378 957
OMVATTEND	NWU/Lapa-korpus 1.1	9 804 270	11 639 129
OMVATTEND	PUK/Protea Boekhuis-korpus 2.1	8 022 403	9 255 228
OMVATTEND	RSG-nuuskorpus 2.1	14 829 223	16 160 631
OMVATTEND	Taalkommissie-korpus 1.1	47 321 344	53 622 677
OMVATTEND	Wikipedia- Afrikaanse korpus 1.0	11 523 680	13 119 966
EKSKLUSIEF	Watkykky.co.za-korpus 1.2	1 236 214	1 406 709
	TOTAAL	132 916 578	150 342 166

Alle korpora in Korpusportaal is outomaties gelemmatiseer (omtrent 90% akkuraat) en van woordsoorte voorsien (omtrent 75% akkuraat). Die annotasies is nie gekorrigeer nie; gebruikers moet dus rekening hou met moontlike annotasiefoute.

Gebruik die volgende verwysing om na hierdie korpusversameling te verwys:

Virtuele Instituut vir Afrikaans (VivA). 2016. Korpusportaal. Beschikbaar by: <http://viva-afrikaans.org>.

6.2 Linguistiese annotasie

6.2.1 Etiket

Annotasie	Etiket	CQL-voorbeeld	Voorbeelde
-----------	--------	---------------	------------

Woord	word	"rooi" [word="rooi"]	<i>rooi, Rooi</i>
Lemma	lemma	[lemma="rooi"]	<i>rooi, Rooi, rooie, Rooie</i>
Woordsoort	pos	[pos="B.NW.*"]	<i>toegesneeude, flikkerende, goeie</i>

6.2.2 Woordsoorte algemeen

Woordsoort	Afkorting	CQL soekterm
Byvoeglike naamwoord	B.NW.	[pos="B.NW.*"]
Bywoord	BW.	[pos="BW.*"]
Lidwoord	LID	[pos="LID.*"]
Naamwoord		[pos="S.* EIE.*"]
- Selfstandige naamwoord	S.NW.	[pos="S.NW.*"]
- Eienaam	EIE.	[pos="EIE.*"]
Punktuasie	U.	[pos="U. (sinseinde sinmiddel links.* regs.*)"]
Telwoord	TW.	[pos="TW.*"]
Tussenwerpsel	TSW.	[pos="TSW.*"]
Voegwoord	VG.	[pos="VG.*"]
Voornaamwoord	VNW.	[pos="VNW.*"]
Voorsetsel	VS.	[pos="VS.*"]
Werkwoord	WW.	[pos="WW.*"]

6.2.3 Woordsoorte gedetailleerd

Etiket in Korpusportaal. Opgespoor per woordsoort via Kundig, bv. [pos="S.NW.*"], Trefslae in groepe, gesorteer op 'hit woordsoort'.

Woordsoort	Etiket
Byvoeglike naamwoord	
	B.NW.stellend.attributief
	B.NW.stellend.predikatief
	B.NW.oortreffend.attributief
	B.NW.vergotend.attributief
	B.NW.vergotend.predikatief
	B.NW.oortreffend.predikatief
Bywoord	

	BW.oortreffend
	BW.stellend
	BW.vergrotend
Lidwoord	
	LID.bepaald
	LID.onbepaald
Selfstandige naamwoord	
	S.NW.abstrak
	S.NW.maatnaam.enkelvoud.basis
	S.NW.massanaam
	S.NW.soortnaam.enkelvoud.basis
	S.NW.soortnaam.enkelvoud.diminutief
	S.NW.soortnaam.meervoud.basis
	S.NW.soortnaam.meervoud.diminutief
Eienaam	
	EIE.eienaam.enkelvoud.basis
Punktuasie	
	U.sinseinde
	U.sinmiddel
	U.links-parentese
	U.regs-parentese
Partikel	
	U.partikel.infinitief
	U.partikel.ontkenning
	U.partikel.ww.
	U.partikel.genitief
	U.partikel.vergelyking
	U.partikel.deel
	U.partikel.graad
	U.partikel.betreklik
Uniek	
	U.eks-daar
	U.woorddeel
Residu	
	R.afkorting

	R.akroniem.letterklankwoord
	R.akroniem.letternaamwoord
	R.ongeklassificeerd
	R.simbool
	R.vreemdetaalwoord
Telwoord	
	TW.hooftelwoord.adjektief.bepaald
	TW.hooftelwoord.adjektief.onbepaald
	TW.hooftelwoord.bywoord.bepaald
	TW.hooftelwoord.bywoord.onbepaald
	TW.hooftelwoord.voornaamwoord.bepaald
	TW.hooftelwoord.voornaamwoord.onbepaald
	TW.rangtelwoord.adjektief.bepaald
	TW.rangtelwoord.adjektief.onbepaald
	TW.rangtelwoord.bywoord.bepaald
	TW.rangtelwoord.bywoord.onbepaald
Voegwoord	
	VG.neweskikkend
	VG.onderskikkend
Voornaamwoord	
	VNW.aanwysend
	VNW.betreklik
	VNW.derde.manlik.enkelvoud.besitlik
	VNW.derde.manlik.enkelvoud.gemarkeerd.persoonlik
	VNW.derde.manlik.enkelvoud.ongemarkeerd.persoonlik
	VNW.derde.manlik.enkelvoud.wederkerend
	VNW.derde.meervoud.besitlik
	VNW.derde.meervoud.persoonlik
	VNW.derde.meervoud.wederkerend
	VNW.derde.onsydig.enkelvoud.ongemarkeerd.persoonlik
	VNW.derde.onsydig.enkelvoud.wederkerend
	VNW.derde.vroulik.enkelvoud.besitlik
	VNW.derde.vroulik.enkelvoud.gemarkeerd.persoonlik
	VNW.derde.vroulik.enkelvoud.ongemarkeerd.persoonlik
	VNW.eerste.enkelvoud.besitlik

	VNW.eerste.enkelvoud.gemarkeerd.persoonlik
	VNW.eerste.enkelvoud.ongemarkeerd.persoonlik
	VNW.eerste.meervoud.besitlik
	VNW.eerste.meervoud.persoonlik
	VNW.eerste.meervoud.wederkerend
	VNW.onbepaald
	VNW.tweede.enkelvoud.besitlik
	VNW.tweede.enkelvoud.gemarkeerd.persoonlik
	VNW.tweede.enkelvoud.ongemarkeerd.persoonlik
	VNW.tweede.enkelvoud.wederkerend
	VNW.tweede.meervoud.persoonlik
	VNW.vraend
	VNW.wederkerig
Voorsetsel	
	VS.voorsetsel
Werkwoord	
	WW.gemarkeerd.hoof.onskeibaar.koppel
	WW.gemarkeerd.hoof.onskeibaar.onoorganklik
	WW.gemarkeerd.hoof.onskeibaar.oorganklik
	WW.hulp.onskeibaar.tyd
	WW.ongemarkeerd.hoof.onskeibaar.koppel
	WW.ongemarkeerd.hoof.onskeibaar.onoorganklik
	WW.ongemarkeerd.hoof.onskeibaar.oorganklik
	WW.ongemarkeerd.hoof.onskeibaar.voorsetsel
	WW.ongemarkeerd.hoof.skeibaar.onoorganklik
	WW.ongemarkeerd.hoof.skeibaar.oorganklik
	WW.teenwoordig.hulp.onskeibaar.aspek
	WW.teenwoordig.hulp.onskeibaar.modaliteit
	WW.teenwoordig.hulp.onskeibaar.modus
	WW.verlede.hulp.onskeibaar.modaliteit
	WW.verlede.hulp.onskeibaar.modus
	WW.gemarkeerd.hoof.onskeibaar.koppel
	WW.gemarkeerd.hoof.onskeibaar.onoorganklik
	WW.gemarkeerd.hoof.onskeibaar.oorganklik
	WW.hulp.onskeibaar.tyd

	WW.ongemarkeerd.hoof.onskeibaar.koppel
	WW.ongemarkeerd.hoof.onskeibaar.onoorganklik
	WW.ongemarkeerd.hoof.onskeibaar.oorganklik
	WW.ongemarkeerd.hoof.onskeibaar.vorsetsel
	WW.ongemarkeerd.hoof.skeibaar.onoorganklik
	WW.ongemarkeerd.hoof.skeibaar.oorganklik
	WW.teenwoordig.hulp.onskeibaar.aspek
	WW.teenwoordig.hulp.onskeibaar.modaliteit
	WW.teenwoordig.hulp.onskeibaar.modus

6.2.4 Boolese operatore

Operator	Simbool	Voorbeeld	Enkele resultate
OR ("of")		"die hierdie daardie" "hond"	<i>die hond, daardie hond, Die hond, Hierdie hond</i>
AND ("en")	&	[lemma="sy" & pos="VNW.*"]	<i>langs sy bevrore lyk, Tom draai weer na sy papier</i>
NOT ("nie")	!	[lemma="assos.*" & pos!="S.*"]	<i>assosieer, geassosieer, Outomobiel Assosiasie</i>

6.2.5 Regex-operatore

Soekopdrag	Invoer	Voorbeeld	Enkele resultate
Willekeurige teken	.	"ma."	<i>man, mal, mat, mag</i>
Karakterklasse	[]	["b[aie]"]	<i>bal, bil, bel</i>
Disjunksie		"ouma oupa"	<i>Ouma, oupa</i>
Groepering	()	["ou(m)p)a"]	<i>Ouma, oupa</i>
Kwantifikasie: 0 of 1	?	["voorbeeld?"]	<i>voorbeeld, voorbeelde</i>

Kwantifikasie: 1 of meer	+	["do+r"]	<i>dor, door door</i>
Kwantifikasie: 0, 1 of meer	*	["do*r"]	<i>dr, dor, door, door</i>
Onbepaald aantal willekeurige tekens (nul ingesluit)	.*	"assos.*"	<i>assosieer, assosiasie, assosiaatdirekteur</i>
Bepaalde aantal	{aantal}	"do{2}r"	<i>door</i>
Aantalinterval	{minimumaantal, maksimumaantal}	"do{1,2}r"	<i>dor, door</i>
Minimumaantal	{minimumaantal,}	"hallo{2,}"	<i>halloo, hallooo</i>
Ontsnapsimbool	\		
Kassensitief	(?-i)	[lemma="(?-i)Man"]	<i>Man, Mans, Manne, Mannetjies</i>

6.2.6 Kwantifikasie van soektogomgewing

Operator	Simbool	Voorbeeld	Enkele resultate
Willekeurig woord	[]	"die" [] "vrou"	<i>die eerste vrou, dié 24-jarige vrou,...</i>
Bepaalde aantal	{aantal}	"baie" []{3} "huise"	<i>baie van die groot huise, baie bly dat hulle huise</i>
Interval	{minimum, maksimum}	[pos="WW.*"]{2,3}	<i>kan help, geplant het, ontdek kan word, gaan sien het</i>
Minimum	{minimumaantal,}	"die" [pos="B.NW.*"]{2,} "man"	<i>Dié merkwaardige jong man, die ware testosteroon-belaaide Afrikaanse man</i>