

DAT ZOEKEN WE OP

Haagse Feiten

In het vorige nummer van IP stond een vraaggesprek met de maker van het parlementaire informatieportaal 'Haagse Feiten'. Het zoekstelsel daarvan bleek interessant genoeg om onze zoekrubriek hiervoor nog even nieuw leven in te blazen.



Door: Eric Sieverts

Haagse Feiten kent een gratis en een betaalde versie. Ook voor die gratis versie moet je je de eerste keer registreren (wat niet erg soepel verloopt). Voor deze test is alleen de gratis versie bekeken, omdat die al alle basiszoekfunctionaliteit biedt. Alleen diverse gespecialiseerde diensten die toch al buiten het kader van deze rubriek vallen, ontbreken daarin. Voor uitgebreider informatie over de bronnen waar Haagse Feiten zijn informatie uithaalt, verwijs ik graag naar het vorige nummer van IP.

Volledig Booleaans

De zoekfunctie bestaat uit een enkele zoekregel om je zoekwoorden in te tikken. De HELP-functie leert dat daarbij ook OR en NOT wordt ondersteund. Booleaanse mogelijkheden zijn echter een stuk vollediger dan de HELP suggereert. Tussen ingetikte zoekwoorden wordt standaard een AND-relatie gelegd, maar die operator mag je er desgewenst wel zelf tussen zetten. OR-relaties werken goed en ook geneste combinaties van AND en OR met haakjes werken foutloos, ook wanneer je daar NOT-operatoren in verwerkt. Hoewel de HELP zegt dat je daarvoor het minteken moet gebruiken, blijkt het woordje NOT ook geaccepteerd te worden. Met aanhalingstekens zoek je zoals gebruikelijk op vaste zinnen. Standaard wordt op woordstammen ge-

zoekt (enkel/meervoud en nog wat meer varianten). De aanhalingstekens werken helaas niet om woordstamzoecken te voorkomen. Met een * kun je ook nog echte truncatie afdwingen.

Filters

Haagse Feiten biedt volledig parametrische filtering. Al voor je gezocht hebt, zie je hoeveel items uit de onderscheiden categorieën in het systeem zitten. Na een zoekopdracht geven die cijfers uiteraard aan hoeveel er van het zoekresultaat overblijft. De hoofdcategorieën voor de parameters waarop je kunt filteren, zijn:

- > **organisaties:** daarbij wordt het begrip 'organisatie' heel ruim opgevat: behalve de ministeries en beide Kamers ook diverse media, zij het dat die laatste wat willekeurig zijn: wel NOS, RTL en NU.nl, maar van kranten alleen *De Volkskrant*.
- > **informatietypen:** daarbij wat ongelijksoortige categorieën als twitter, kamerstuk, nieuws, wetsvoorstel, stemmingsuitslag, bijlage, toezegging, et cetera;
- > **subtypen:** met een nog veel verder uitgewerkte uitsplitsing van het voorgaande;
- > **bronnen:** de sites (op URL) waarvan het materiaal afkomstig is (met dus nogmaals twitter);
- > **personen:** alle voorkomende (oud-) ministers, kamerleden en dergelijke, ook als ze maar een enkele keer voorkomen;

> **dossiers:** een grote verscheidenheid aan onderwerpsdossiers waarin het opgenomen materiaal is ondergebracht.

Opmerkelijk (maar wel handig) is de overlap tussen hoofdcategorieën. Zo zijn Wikipedia, NOS, *De Volkskrant* en dergelijke zowel een organisatie als een bron en twitter zowel een bron als een informatietype. Plezierig is verder dat je binnen een hoofdcategorie ook op meer keuzes kunt filteren.

Waar in je zoekt

Het systeem zoekt in volledige teksten. Daarbij hoort zelfs de hele inhoud van achterliggende pdf's (die je overigens eerst moet downloaden om die inhoud te zien te krijgen). Dit betekent dat je vaak resultaten vindt waarin een zoekterm alleen terloops ergens in een lange opsomming of bijlage voorkomt. Het helpt dan om de sorteervolgorde van datum (wat voor politieke vragen vaak nuttig is) te wijzigen in relevantie ('score' genoemd). In HELP-schermen wordt niet over zoekvelden gerept. Uitproberen of er toch ongedocumenteerde mogelijkheden waren, leerde dat in elk geval op titelwoorden ingeperkt kan worden. Met *titel:vredespaleis* vind je alleen de documenten waar dat zoekwoord in het titelveld voorkomt. Dat biedt dus nog een andere mogelijkheid om de relevantie van je zoekresultaat (sterk) te verhogen. Ook in Booleaanse combinaties en bij exacte phrases kan dit gebruikt worden. De van allerlei verschillende sites gehaalde inhoud van Haagse Feiten zit in een database die zelf niet nog een keer in Google terecht komt. Maar met de zoekmogelijkheden van Haagse Feiten voel je ook nauwelijks behoefte om Google-functionaliteit te willen toepassen. <

URL	haagsefeiten.nl
Booleaans combineren	AND, OR, NOT (-), genest met haakjes
Truncatie	ja (plus ook automatische woordstammen)
Speciale zoekvelden	titel
Advanced zoekscherm	nee
Parametrische filters	ja (grote verscheidenheid)
Ook in Google	nee
Semantische markup	nvt

Eric Sieverts is redacteur van IP en freelance docent en adviseur.