

Citatie-zoeken op het Web

Eric Sieverts

In vorige afleveringen in deze serie werd tot dusverre steeds uitgegaan van een bepaalde zoekmachine en werd verteld wat die allemaal kon. Dit keer bekijken we de zaak eens van de andere kant: we willen iets; welke zoekmachine kan dat? Oftewel: welke zoekmachine kiezen we bij een bepaald soort vraag?

Het onderwerp is deze keer citatie-zoeken. Wie daarbij denkt aan de mooie, maar dure producten van ISI, zoals de Science Citation Index, moet ik een beetje teleurstellen. Een zekere vorm van 'citatie-zoeken' is mogelijk, maar niet op de geavanceerde en kwalitatief redelijk gecontroleerde manier van ISI. Anderzijds bieden een paar zoekmachines wel degelijk interessante mogelijkheden - en het is gratis!

De uitspraak 'een zekere vorm van' klinkt misschien wat vaag. Daarom eerst een beschrijving (of definitie zo u wilt) van wat ik hier in WWW-kader mee bedoel. 'Citeren' is op het Web namelijk wel wat anders dan in de (wetenschappelijke) literatuur, maar ook weer niet zo heel veel anders. Een wetenschappelijk artikel wordt altijd keurig afgesloten met een degelijke lijst literatuurreferenties (het ontbreken van zo'n lijst aan het eind van dit artikel leert u dus meteen dat het geen wetenschappelijke pretenties heeft). Dat zijn de 'citaties' waar ISI het over heeft. Via hun citatie-indexen kun je dan opzoeken wie recent het belangrijke standaardartikel van professor Soulsqueezer over *myalgocephalische effecten in de hypopotamos* heeft aangehaald (geciteerd). Je weet dan namelijk zeker dat die nieuwe artikelen ook weer iets met zulke *myalgocephalische effecten* te maken hebben.

Wat zijn nu citaties op het Web? Daar

hebben we meestal niet te maken met keurige artikelen (al zijn ze er wel) en met keurige lijstjes met (echte) literatuurreferenties (al zijn die er ook wel). In elk geval kunnen de ons bekende zoekmachines die artikelen niet als een apart te behandelen categorie uit het Web vissen. Eigenlijk gaat citeren op het Web veel eenvoudiger, vaak veel uitgebreider en misschien wat ongemerkter. Elke hyperlink in een HTML-document is namelijk een verwijzing en veel van die hyperlinks verwijzen naar andere webdocumenten; dat zijn de links die van het Web een echt web maken, of om een bekend hypertext-boek te parafaseren: *they are the glue that keeps the Web together*. Zo werken hyperlinks dus toch een beetje als de literatuurreferenties in gewone artikelen.

Volgende vragen zijn dan natuurlijk: Heb je er ook wat aan daarop te willen zoeken? En als je het wilt, kan het dan ook? U begrijpt het al: mijn antwoord op beide vragen luidt: 'Ja!'

Waarom zou je het doen?

Net als in de literatuur geldt dat een

webpagina die naar een andere webpagina verwijst, daar inhoudelijk iets mee te maken heeft. Heb je een mooie medische website te pakken, dan zullen webpagina's die daarna verwijzen vast ook over medische onderwerpen gaan. En dat niet alleen: verwijzingen worden meestal alleen maar opgenomen als de maker van een pagina ze voor zichzelf of zijn doelgroep interessant vindt. Dit betekent dat niveau en aard van citerende pagina's niet al te veel zullen verschillen van die van de geciteerde. Je hebt dus goede kans op een zoekresultaat met een veel grotere precisie dan wanneer je wat gewone zoektermen op een zoekmachine loslaat. Je moet dan maar op de koop toenemen dat het ook wel eens pagina's oplevert die alleen maar een uitgebreide lijst verwijzingen bevatten en geen eigen informatie bieden. (Overigens heeft zelfs dat zijn pendant in de wetenschappelijke literatuur, waar soms overzichtsartikelen verschijnen die eigenlijk evenmin eigen informatie toevoegen.)

Hoe werkt het?

Hoe gaat het zoeken hiernaar nu in zijn werk? In feite is het een bijzondere vorm van het 'in velden zoeken' dat een paar zoekmachines bieden. In HTML-documenten zijn namelijk wel degelijk een soort velden te vinden, al zien ze er wat anders uit dan de velden die we in echte (bibliografische) databases gewend zijn. Het zijn delen van documenten die met bepaalde HTML-tags (voor en achter het betreffende tekstfragment) gemarkeerd zijn, bijvoorbeeld `<title>xxx</title>` voor de in de blauwe bovenbalk van Netscape verschijnende titel van een webdocument,

URL's van genoemde zoekmachines

AltaVista: <http://www.altavista.digital.com>

<http://www.altavista.telia.com>

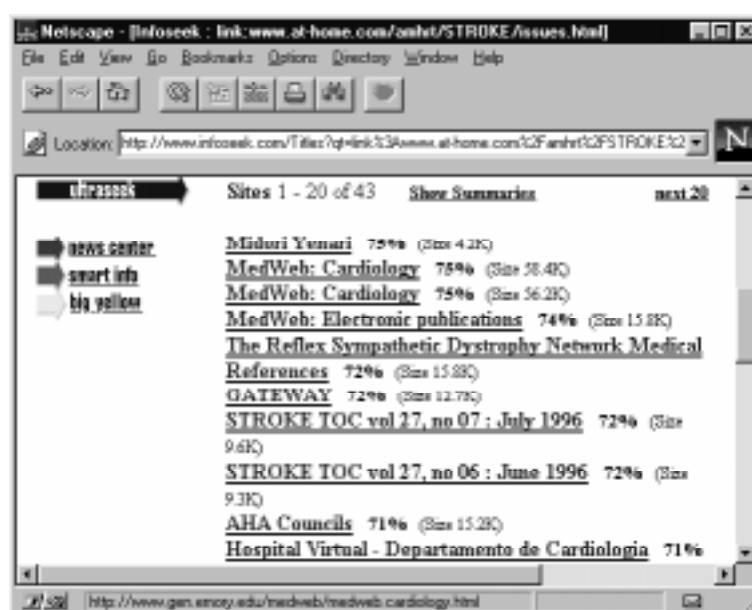
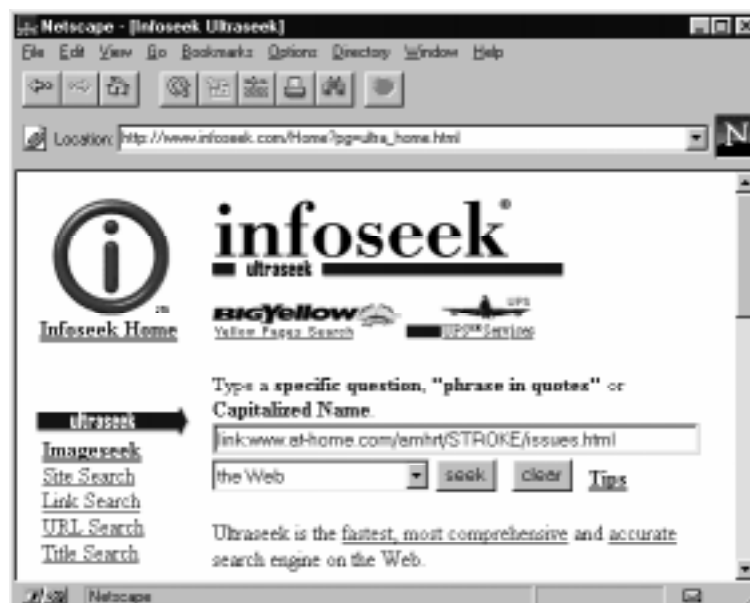
HotBot: <http://www.hotbot.com>

InfoSeek: http://www.infoseek.com/Home?pg=ultra_home.html

<H1>xxx</H1> voor een vette grote kop in de tekst of <AHREF 'URL'>xxx voor een hyperlink. Daarmee zijn we er dan meteen, want zowel InfoSeek als AltaVista bieden de mogelijkheid om in de zoekvraag een veldaanduiding voor dat laatste type veld op te geven. Het intikken van bijvoorbeeld de zoekopdracht `link:www.ubu.ruu.nl` levert dan alle webpagina's die een link hebben naar enige pagina op de WWW-server van de Utrechtse Universiteitsbibliotheek. Je kunt natuurlijk ook wat specifiekere zijn en een volledig URL opgeven, bijvoorbeeld `link:www.at-home.com/amhrt/STROKE/issues.html`, waarmee gezocht wordt naar alle pagina's die verwijzen naar het online tijdschrift 'Stroke' van de American Heart Association.

Zelf-citaties

Misschien nog wel sterker dan bij literatuurreferenties komen op het Web ook 'zelf-citaties' voor, hier alleen met een wat minder negatieve connotatie dan bij de echte citatie-indexen. In een hypertext-omgeving is het immers heel gebruikelijk en vooral ook functioneel om tussen een groep gerelateerde documenten binnen eenzelfde website heen en weer te linken. Toch zul je ook bij het zoeken op het Web dit soort zelf-citaties graag kwijtraken, omdat je juist wilt weten op welke andere sites informatie over hetzelfde onderwerp te vinden is. Ook daar zijn mogelijkheden voor, als je ook op URL's van te vinden pagina's kunt zoeken, of liever gezegd: als je die bij het zoeken kunt uitsluiten. Een voorbeeld daarvan is de bij InfoSeek te geven zoekopdracht `+link:ubu.ruu.nl -site:ruu.nl`, waarmee pagina's op alle web servers van de Utrechtse Universiteit die naar de bibliotheek verwijzen, van het zoekresultaat worden uitgesloten. (Het veld 'link' bevat dus de URL's die in hyperlinks in een document voorkomen, het veld 'site' de URL van de (niet) te vinden pagina zelf; de + geeft aan dat een term



verplicht moet voorkomen, de - dat een term beslist niet mag voorkomen.) Zoals het voorbeeld laat zien, kun je hierbij dus zelfs fragmenten van Internetadressen gebruiken. Bij AltaVista wordt hetzelfde bereikt met als uit te sluiten veld 'url:' of 'host:', dus bijvoorbeeld: `+link:ubu.ruu.nl -host:ruu.nl`.

Nog een zoekmachine

Op een wat andere manier kun je ook bij HotBot aan citatie-zoeken doen. Daar biedt een uitrolhokje in het zoekscherm keuzes om met de ingetikte zoekterm(en) te zoeken for 'all the words', 'any of the words', 'the exact phrase', 'the person' (daarover later nog eens iets), 'the Boolean expression' of inderdaad 'links

to this URL'. Anders dan bij beide andere zoekmachines kun je hier geen fragmenten van URL's opgeven, maar moet je een volledig URL als zoekterm opgeven, zelfs voorafgegaan door de aanduiding 'http://'. Hoewel HotBot vervolgens wel de mogelijkheid biedt allerlei verdere inperkingen te maken, is uitsluiten van bepaalde adressen om zelfcitaties te elimineren bij HotBot niet mogelijk. Mijn slotadvies: citatie-zoeken op het Web, probeer het eens.

Dr. E.G. Sieverts is werkzaam op de Hogeschool van Amsterdam en bij de Universiteitsbibliotheek Utrecht. Hij is redacteur van Informatie Professional.