

# Google Scholar: reus zonder schouder

Henk Ellermann h.e.ellermann@ub.rug.nl



*Het is in wezen schandig, maar een groot deel van de wetenschappelijke resultaten bevindt zich op het zogeheten invisible web. Wachtwoorden, firewalls en andere ongein voorkomen dat niet betalende of anderszins ongerechtigde gebruikers wetenschappelijke publicaties kunnen inzien.*

Lange tijd is ook voor Google het invisible web onzichtbaar gebleven. Met de komst van Google Scholar lijkt dat te veranderen. Door afspraken te maken met universiteiten, wetenschappelijke uitgevers, beheerders van wetenschapsporthals, wetenschappelijke genootschappen en dergelijke, krijgt Google Scholar wel de mogelijkheid om op het onzichtbare web te zoeken.

De afspraken, zo verzekert Google ons, gaan niet gepaard met financiële transacties. Nota bene, de inhoud zelf blijft ontoegankelijk voor ongerechtigden, maar Google mag het wel gratis indexeren en een link genereren naar de afgeschermdede inhoud. Google Scholar combineert de informatie van het onzichtbare web met die informatie op het gewone zichtbare web waarvan het aanneemt dat de inhoud enige wetenschappelijke waarde heeft, en maakt het doorzoekbaar.

## Page Rank algoritme

De kracht van Google, al sinds haar officiële oprichting in 1998, en nu ook van Google Scholar, is de manier waarop zoekresultaten geordend worden. De belangrijkste pagina's komen bovenaan, waarbij het belang wordt afgemeten aan het aantal links dat een pa-

gina ontvangt, gewogen met het belang van de verwijzende pagina's zelf. De circulariteit die in deze definitie verborgen zit, wordt omzeild met behulp van het gepatenteerde PageRank algoritme.

In het geval van Google Scholar wordt niet alleen gebruik gemaakt van de op een website aanwezige hyperlinks, zoals bij de gewone Google, maar worden ook de teksten zelf geanalyseerd. De auteursnamen en de referenties in de artikelen worden geëxtraheerd en gebruikt om het belang van een artikel te bepalen. Wordt een auteur vaak geciteerd dan krijgt het artikel een aantal bonuspunten en hetzelfde gebeurt als aan het artikel vaker gerefereerd wordt.

Voor mijn vakgebied, de digitale bibliotheek, heb ik steekproefsgewijs gevonden dat de klassiekers op basis van relevante sleutelwoorden inderdaad ergens bovenaan de zoekresultatenlijst te vinden waren. Het is niet duidelijk of voor andere vakgebieden de resultaten even positief zijn en evenmin is duidelijk of eventuele negatieve resultaten veroorzaakt worden door de analysemethode of door een gebrekkige penetratie van het onzichtbare web; Google Scholar zal maar net geen deal hebben kunnen sluiten met een belangrijke uitgever op uw terrein, nietwaar? Voor gedegen antwoorden op zulke vragen moet je weten welke sites Google Scholar doorzoekt en hoe de rangordening precies bepaald wordt. Ter zake relevante informatie wordt door Google Scholar ech-

ter niet verstrekt. Het onzichtbare web wordt zichtbaar gemaakt met een hulpmiddel dat zelf onzichtbaar is.

## Bezwaar

Het is mogelijk Google Scholar op vele detailpunten te bekritisieren. U kunt de kritiek ook zelf eenvoudig tussen twee slokken thee door verzinnen. Er zal nooit een degelijk (semi-)automatisch onderscheid gemaakt kunnen worden tussen wat wel en niet wetenschappelijk is, een automatische analyse van gevonden teksten zal minder zijn dan een analyse door deskundige catalogiseerders, ontdebbling van gevonden artikelen en auteursnamen is problematisch, links zijn niet gegarandeerd permanent bruikbaar, en de deuren waarachter Google zelf wel mag kijken blijven soms gesloten voor een groot deel van de gebruikers. Het is allemaal waar, natuurlijk, maar het gaat hier eerder om beperkingen in techniek, om keuzes en uitdagingen dan om heuse kritiek. Hét bezwaar tegen Google Scholar blijft, dat men niet open is over hun werkwijze: je weet niet (precies) wat je ziet.

Het zelfportret van Google Scholar is dat ze staat op 'the shoulders of giants' en graag een dienst wil teruggeven aan de wetenschap waaraan het zelf zoveel te danken heeft. Zolang men echter met onzichtbare gereedschappen werkt, biedt ze zelf geen schouder aan om op te staan. Google Scholar is een reus zonder schouder. <

## Links

- Google Scholar: [scholar.google.com](http://scholar.google.com)
- Henk Ellermann is hoofd digitale bibliotheekvoorzieningen van de Universiteitsbibliotheek Groningen