

## De mythe van Google

Geschreven door Eric Sieverts  
dinsdag, 30 juni 2009

---

[Zie ook ["De mythe van de catalogus"](#) ]

Marketeers leren ons dat we daar moeten zijn waar onze gebruikers zijn. Onze gebruikers "zijn op Google". Dus zouden wij ook "op Google" moeten zijn. Maar wat bedoelen we daar eigenlijk mee? Vaak hoor ik dat we zouden moeten zorgen dat de boekbeschrijvingen uit onze catalogi "in Google" zitten. Dan komen onze klanten daar via Google wel terecht. Maar helpt dat?



Laten we eens een rekensommetje maken, waarvoor we voor het gemak alle Nederlandse UB's maar samen nemen.

Die Nederlandse UB's zullen samen zo'n 5 miljoen (?) titels hebben. Daartegenover heeft Google vermoedelijk al meer dan 500 miljard "titels".

Als je een willekeurige zoekactie doet, zal dus maar 1 op elke 100.000 resultaten een boek uit een UB zijn.

Maar het is nog erger. Google zoekt full-text in webpagina's, PDF's, Word-documenten enzovoort. Een gemiddeld "document" dat Google indexeert zal dus zeker 1000 woorden bevatten en waarschijnlijk nog aanzienlijk meer. De "documenten" die wij in Google zouden moeten krijgen, zijn echter catalogusrecords.

Dat wil zeggen metadata, waar we ontzettend trots op zijn dat we die zo consistent en uniform en gestandaardiseerd hebben weten in te voeren, maar die, zoals ik in mijn "[mythe van de catalogus](#)" al aangaf, niet meer doorzoekbare tekst bevatten dan alleen een titel van pakweg 10 woorden, plus nog hooguit 5 trefwoorden.

Dat is maar 1/60 van het aantal woorden (dus zoekingen) dat een gemiddeld Google document heeft.

Zo verder rekenend, zou dus maar 1 op de 6 miljoen zoekresultaten een catalogusrecord van een boek uit onze collecties zijn. Denken we echt dat onze boeken dan op Google gevonden worden?

Toegegeven, mijn berekening van deze vindkans van 1 op 6.000.000 is een beetje eenzijdig.

## De mythe van Google

Geschreven door Eric Sieverts  
dinsdag, 30 juni 2009

---

In de eerste plaats geldt deze berekening voor onderwerpszoekacties, niet voor een "known item search". Als je al een auteursnaam en/of (ongeveer) een titel weet, ligt de zaak ietsje anders, maar ook dan zullen de meeste catalogusrecords nog steeds verdrinken in de overmaat aan andere resultaten uit Google.

Verder hield mijn rekensommetje nog geen rekening met de relevance ranking die zoekmachines als Google toepassen. Het nadeel van onze magere metadata verkeert daarbij ineens nog in een voordeel. Doordat catalogusrecords nauwelijks meer ingangen hebben dan alleen de titelwoorden, komen de zoekwoorden waarop onze boeken gevonden worden, als ze gevonden worden, vrijwel altijd in die titels voor. En bij vrijwel alle zoekmachines is dat een van de belangrijkste criteria voor een hoge ranking. Van de webpagina's die Google op een zoekvraag vindt, zal een belangrijk deel dus lager scoren dan die catalogusbeschrijving, omdat de zoekwoorden in veel van die webpagina's pas verderop in hun veel langere teksten staan. Bovendien zijn onze bibliotheken veelal gerespecteerde instellingen, ook in termen van Google's populariteitsmeting, de pagerank.

Maar als we echt hoog willen scoren zullen we met onze catalogi ook nog aan Search Engine Optimization moeten doen. En dat is een techniek die ik eerder in [een column](#) al eens met doping heb vergeleken. Als iedereen optimaal optimaliseert, wat steeds meer het geval lijkt, dan schiet je daar ook niets meer mee op. Maar je kunt het ook niet meer achterwege laten, omdat de anderen het allemaal wel blijven doen en dan dus harder zouden fietsen dan jij, ook al zijn ze niet echt beter.

Waarschijnlijk ligt het dus genuanceerder, maar zelfs bij een vindkans van 1 op 6.000 - drie ordes van grootte gunstiger dan mijn aanvankelijke berekening opleverde - blijft het feit dat onze boeken op onderwerpszoekacties in Google door onze klanten veel vaker niet dan wel gevonden zullen worden. En dat ligt dus niet alleen aan de overmaat aan andere informatie in Google, maar ook aan onze eigen kale catalogi, waarin boeken vaak helemaal niet te vinden zijn op de onderwerpen die daarin aan de orde komen. Dat laatste is een stokpaardje dat ik onder het kopje "[de mythe van de catalogus](#)" al apart heb bereden.

Dan maar helemaal geen pogingen meer doen om "in Google" te komen?

Ik betwijfel inderdaad of het zinnig is dat zelf te willen. Laten we dat liever overlaten aan grote spelers als Worldcat. In zo'n gemeenschappelijke catalogus zullen vaak al meer metadata beschikbaar zijn, dan in ieders lokale catalogus (zoals ook in enkele van de reacties op mijn vorige tekst al gezegd werd). Daar zijn bovendien al diensten beschikbaar om te zien te krijgen welke gevonden boeken eventueel in naburige bibliotheken te vinden zijn. En - misschien wel het belangrijkste - daarbij is het vooralsnog makkelijker om gebruik te maken van extra mash-ups die (bijvoorbeeld via ISBN's) mogelijk zijn met resultaten uit Google of Google Books, zodat gebruikers via omwegen (toevallig) misschien toch nog bij je eigen fysieke collectie terecht komen.