

Van pagina's zoeken naar antwoorden geven

Google vierde onlangs zijn vijftiende verjaardag. Een aanleiding om stil te staan bij wat de zoekgigant in die vijftien jaar teweeg heeft gebracht.

Eric Sieverts

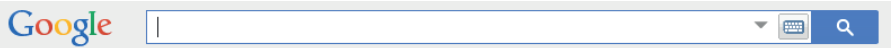
Vele ambities

Googles merknaam is de algemene aanduiding voor zoeken geworden; Google dat maar even. Dat bewijst dat ze het zoeken van informatie binnen ieders belangstelling hebben gebracht. Dat is niet langer aan informatiespecialisten voorbehouden. De oorspronkelijke ambitie alle kennis van deze wereld vindbaar te willen maken, lijkt de laatste tijd wat verwaterd, nu ze zich hebben ontwikkeld tot het grootste advertentiebedrijf ter wereld, waarbij de zoekmachine nog slechts een hulpmiddel is om advertenties te verkopen.

Niettemin hebben ze intussen een grote diversiteit aan zoekmachines ontwikkeld, voor allerlei soorten media en materialen, zelfs allemaal gecombineerd in Googles Universal Search. De vraag dringt zich dan op of die ook allemaal apart zullen blijven bestaan. In het bijzonder maken wetenschappelijke zoekers zich zorgen over de toekomst van Google Scholar. De waslijst aan verdwenen diensten en producten geeft wat dat betreft reden tot ongerustheid. Daarin openbaart zich het gevaar van Googles machtspositie – er is een bijna-monopolie op zoeken ontstaan.

Daarnaast is intussen een veelheid aan diensten en producten ontwikkeld (of opgekocht), die niet rechtstreeks met zoeken te maken hebben. Sommige nog wel met informatievoorziening in het algemeen, zoals Blogger, Gmail, Google Maps, Google+, Google Drive en zelfs Google Glass. Maar soms ook heel buitenissige, zoals de zelfstandig rijdende Google-auto.

Over nog eens vijftien jaar zullen we weten wat er dan van alle ambities en producten geworden is.



Het eerste jaar

De traditie wil dat Google, zoals veel technologiebedrijven, in een garage is ontstaan. Toch is dat niet helemaal waar, want Larry Page en Sergey Brin hadden in 1998 hun nieuwe zoekmachine ontwikkeld als onderdeel van hun studie aan Stanford University. Het allereerste url was dan ook <http://google.stanford.edu>. Pas toen Google echt een product werd, kwam die garage in beeld. Toen hadden ze intussen al 25 miljoen pagina's geïndexeerd.

Het nieuwe aan hun systeem was, dat relevantie van gevonden pagina's niet meer alleen werd bepaald door aanwezigheid en locatie van zoektermen, maar ook door iets als kwaliteit. Geïnspireerd door de citatieanalyses van wetenschappelijke publicaties, bedachten ze een slim algoritme dat linkpatronen tussen webpagina's analyseerde en daaruit een kwaliteitsmaat afleidde. Dat was de Pagerank, die voor elke pagina in een iteratief proces werd uitgerekend. Hoe meer links van andere pagina's, en hoe hogere pagerank die zelf hebben, hoe beter. Daarbij heeft de 'page' in 'pagerank' niet alleen betrekking op die 'pagina's', maar is het ook een toespeling op Larry's achternaam.

Dat Google daarna zo snel populair werd, kwam door een combinatie van factoren:



- > de eenvoudige kale interface (dat een benchmark voor zoekinterfaces werd),
- > een zorgvuldig opgebouwd imago van 'no evil',
- > en uiteraard veel betere zoekresultaten.

Zoekmachines als AltaVista en ExCite bleken namelijk steeds minder opgewassen tegen commerciële zoekwoordspamming. Ontevreden zoekers weken daardoor steeds meer uit naar onderwerpsgidsen als Yahoo- en Open Directory. Google heeft dat tij helemaal doen keren.

Kerkhof van verdwenen diensten

Google was altijd sterk in het ontwikkelen van nieuwe functionaliteit en diensten. Vaak waren ze minder goed daar voldoende publiciteit voor te maken, waardoor veel ervan maar weinig werd gebruikt. En dat was dan een reden om die functie of dienst na een tijdje – al dan niet geruisloos – te laten verdwijnen. Zo is een omvangrijk kerkhof met verdwenen functies, diensten en producten ontstaan. In verborgen hoekjes liggen allerlei producten die langzaam waren weggekijnd en waarvan bijna niemand zich meer het bestaan herinnert. Maar langs de hoofdwegen liggen ook functies en producten waarvan door velen wordt betreurd dat Google ze meedogenloos heeft omgebracht. Een ware ‘graveyard of broken dreams’.

Hier een rijtje van verdwenen functies en diensten die ik zelf, soms intensief, soms incidenteel, en soms zelfs helemaal nooit gebruikt heb: +-operator (exact search), ~-operator (synonym search), Translated search, Desktop search, Google Wave, Google Buzz, Google Reader, iGoogle, Google Directory, Google Gears, Google Notebook, Google Answers, Google Health, Google Toolbar, Google Pack, Google Aardvark, Google Talk, Google Labs, Google Sets, Code search, Sidewiki, Searchwiki, Wonderwheel, Timeline, Related Links, Knol, Image Swirl, ...

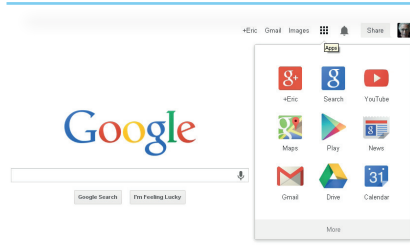
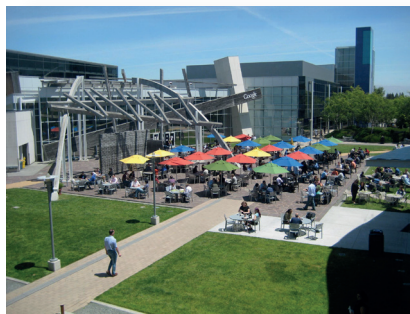
Indrukwekkend? Phil Bradley somde er vorig jaar in zijn weblog (goo.gl/6Qs9qt) nog tientallen meer op.

Google nu

Anno 2013 is Google een bedrijf van grote getallen geworden. De hoofdvesting, het Googleplex, is een campus van 11 hectare, waar volgend jaar nog 10 hectare bijkomt. Een aardige groei vergeleken met de garage uit 1998. Dat geldt ook voor het aantal doorzoekbare pagina's. Google geeft daar zelf geen cijfers over, maar men schat dat dat nu in de orde van 500 miljard ligt, 20.000 maal zo veel als in 1998. Wat Google wel meldt, is dat ze 60 biljoen verschillende url's kennen. Een 6 met 13 nullen. Daarvan is dus maar nauwelijks 1 procent ook werkelijk geïndexeerd. Evengoed is die index al 100 miljoen gigabyte. Daarop worden 4 miljoen zoekacties per minuut uitgevoerd (tegen 7 zoekacties in 1998). Daarmee ligt Googles marktaandeel op ongeveer 75 procent – in Nederland zelfs op 97 procent.

Google heeft serverparken over de hele wereld, met een opslagcapaciteit van enkele exabytes. Hun energiegebruik is dat van een flinke provinciestad. Met een jaaromzet van 55 miljard dollar en een winst van 12 miljard dollar, hoort Google toch nog lang niet tot de echt grote bedrijven.

De Pagerank die Google groot maakte, heeft recent aan belang ingeboet. Midtels Googles Knowledge Graph en door



toepassing van semantische technieken, zoals de Schema.org-ontologie (ook ondersteund door Bing en Yahoo), evolueert Google naar een ‘antwoordmachine’. Daarbij hebben ze geen last van gemanipuleerde metatags, omdat het niet meer om de inhoud van trefwoorden gaat, maar om codering van de aard van woorden en getallen in gevonden documenten. Intussen geeft Google Now zelfs al antwoord op vragen die nog niet gesteld zijn.

Eric Sieverts is redacteur van InformatieProfessional.

Google-generatie aan het woord

‘Toen Google geboren werd, was ik zeven jaar. Ik behoor dus echt tot de Google-generatie die nooit anders gekend heeft. Vanaf mijn scholientijd lag de informatiewereld als een groot Google-boek aan mijn voeten. Toen ik een werkstuk maakte over het oude Egypte, keek ik niet langer in de Winkler Prins of snuffelde ik door boeken. Ik zocht op Google, daar stond actuele informatie en ik vond er indrukwekkende afbeeldingen. Bij een studieproject op de HvA brainstormde ik thuis via Google Drive (eerder Google Docs) met andere projectleden, waardoor wijzigingen realtime werden bijgewerkt en we niet voor één uur overleg naar school hoefden te reizen. Toen de spionagepraktijken van de NSA aan het licht kwamen, heeft me dat niet geremd in het gebruik van Googles gebruiksvriendelijke diensten, omdat er geen vergelijkbare alternatieven zijn. Met dat in het achterhoofd vind ik het prima dat Google – tot op zekere hoogte – geld met mijn persoonlijke gegevens verdient om zo de kosten te dekken.’

Tom de Munck studeert Media, Informatie en Communicatie aan de HvA en loopt nu stage bij Dutch View in Hilversum.