

# 5. DE BETEKENIS VAN ZOEKMACHINES IN DE INFORMATIESAMENLEVING

DOOR PROF. DR. NICÓ VAN EIJK

Voor het vinden van informatie wordt in toenemende mate een beroep gedaan op het internet. Zoekmachines zijn daarbij onontbeerlijk en vormen de belangrijkste schakel tussen degene die informatie aanbiedt en de degene die informatie zoekt. Kortom, zoekmachines vervullen door het ontsluiten van informatie op het internet een centrale rol bij de kennisverwerving in de moderne informatiemaatschappij. Een belangrijke beperking betreft de vraag of het zoekresultaat wel altijd dat oplevert wat gezocht wordt: zoekresultaten worden volop gemanipuleerd. De hoogste plaats op de lijst met zoekresultaten is te koop en aanbieders van informatie maken gebruik van geavanceerde methodes om zoekmachines te misleiden. Daar komt bij dat met name reclame de belangrijkste bron van inkomsten is en de maximalisering ervan een belangrijke, zo niet de belangrijkste, 'driver' van het onderliggende business model is: staat het bieden van het zoekresultaat centraal of gaat het om het optimaliseren van de inkomsten?

## De zoekmachine is onmisbaar geworden

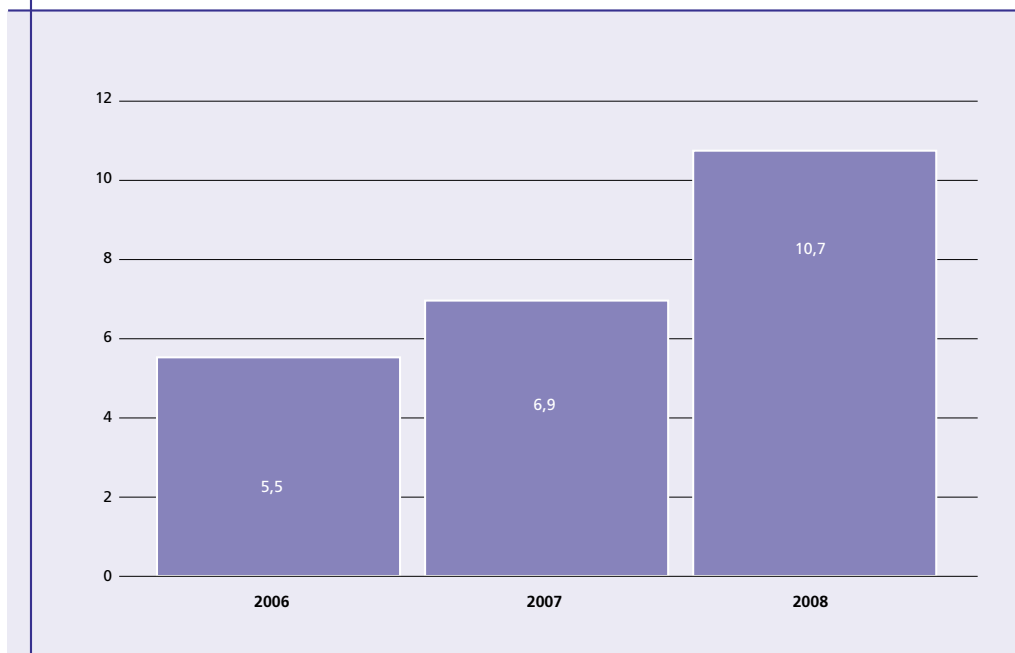
Zoekmachines zijn onmisbaar geworden bij het ontsluiten van digitale informatie. Ze worden gebruikt door vrijwel alle internetgebruikers, die bovendien van mening zijn dat zoeken via zoekmachines de beste manier is om websites te vinden. Het 'Google'-en is een zelfstandig begrip geworden en een eigen vorm van vrijetijdsbesteding, vergelijkbaar met het zappen in het aanbod op televisie. Wie niet via een zoekmachine is te vinden, bestaat niet. 'To exist is to be indexed by a search engine' is in dit verband een veel aangehaald citaat.<sup>1</sup>

Met name Amerikaans onderzoek biedt interessante gegevens over het vinden van informatie op het internet. Regelmatig doet het Pew Internet & American Life Project onderzoek naar de betekenis van informatievoorziening via het internet. Een van de onderzoeken ('Information searches that solve problems') geeft onder meer aan dat bij het zoeken van hulp bij het oplossen van problemen/vragen (bijvoorbeeld over gezondheid, overheid, onderwijs) zo'n 40 tot 60 procent gebruik maakt van het internet als bron. In vrijwel alle gevallen behoort het internet tot de top-3 van bronnen (naast bijvoorbeeld vrienden/kennissen en professionele hulpverleners). Een ander onderzoek over gezondheidsinformatie ('Online Health Search 2006') stelt dat 66 procent van de informatiezoekers de laatste keer dat zij informatie zochten, gebruik maakten van een zoekmachine (27 procent ging rechtstreeks naar een relevante site). Meer dan de helft van de respondenten heeft wat met de gevonden informatie gedaan. Maar tegelijkertijd wijst het onderzoek ook uit dat slechts een kwart van de informatiezoekers de gevonden informatie altijd checkt. Een derde onderzoeksrapport ('Search Engine Users'), kreeg als pakkende onderkop: 'Internet searchers are confident, satisfied and trusting – but they are also unaware and naïve'. Hoewel het betreffende rapport in 2005 is gemaakt – en we misschien mogen aannemen dat internetgebruikers inmiddels wat minder naïef zijn – laat het toch zien dat gebruikers een zeer groot vertrouwen in zoekmachines hebben: 68 procent van de zoekmachinegebruikers zeggen dat zoekmachines een objectieve en onpartijdige bron voor informatie zijn ('a fair and unbiased source of information'). Slechts 19 procent wantrouwt zoekmachines. Overigens geeft weer ander recent onderzoek aan dat gebruikers zoekresultaten meer relevant achten naar mate ze hoger in de lijst met zoekresultaten terecht komen (een voor de hand liggende conclusie, maar toch mooi dat die met onderzoek wordt onderbouwd). Het onderzoek uit 2005 laat ook zien dat er buitengewoon veel gebruik wordt gemaakt van zoekmachines: 84 procent van de internetgebruikers maakt gebruik van zoekmachines en 58 procent doet dat dagelijks. Een verdere illustratie van het omvangrijke gebruik van zoekmachines is te ontleen aan de cijfers over het aantal zoekopdrachten in de Verenigde Staten (figuur 5.1).

Er zijn geen directe redenen om te veronderstellen dat de trends die uit deze Amerikaanse gegevens zijn af te leiden, zich niet op vergelijkbare wijze in Europa of in Nederland voordoen. Volgens de 'Nationale Search Engine monitor' begint maar liefst 62 procent van de Nederlandse

FIGUUR 5.1

AANTAL ZOEKOPDRACHTEN IN DE VERENIGDE STATEN (X MILJARD)



Bron data: ComScore

internetgebruikers zijn zoektocht met een zoekmachine. Van de gebruikers van Google vindt 52 procent bijna altijd wat men zocht (of denkt dat). Google staat ook nummer één in een lijstje van websites waar de meeste tijd aan wordt besteed (tabel 5.1). In maart 2008 spendeerden Nederlandse internetgebruikers niet minder dan 38,5 miljoen uur op Google.

TABEL 5.1

TOP-TIEN WEBSITES NAAR TIJDBESTEDING IN 2008 IN NEDERLAND PER MAAND

WEBSITE GERANGSCHIKT NAAR TIJDBESTEDING	(IN MILJOEN UUR)
Google	38,5
Hyves	19,6
NU	10,9
Relatieplanet	8,5
Startpagina	6,8
AD	5,6
MSN	5,5
Zigiz	4,1
Marktplaats	4,1
Youtube (NL)	3,3

Bron data: Multiscope

## De werking van een zoekmachine

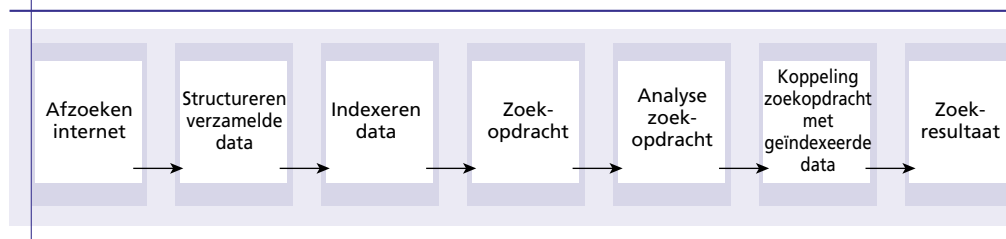
De zoekmachine vervult vooral een toegangsfunctie; het is een gateway naar mogelijk relevante informatie op het internet. Het is echter een toegangspoort in twee richtingen, namelijk van de informatieaanbieder naar de gebruiker en van de gebruiker naar de informatieaanbieder. De zoekmachine bepaalt welke informatie van een informatieaanbieder door de eindgebruiker gevonden kan worden maar ook welke informatie de eindgebruiker uiteindelijk zal vinden. De aangeboden zoekfaciliteit en het onderliggende zoekalgoritme regelen zo het aanbod en de vraag.

Hoe gaat een zoekmachine daarbij te werk? Voor de meeste zoekmachines geldt dat zij *grosso modo* eenzelfde methodiek volgen om tot zoekresultaten te komen. Het proces start met het afzoeken van het internet naar informatie. Dit geautomatiseerde proces gebeurt door intelligente 'speurders', ook wel *spiders*, *bots* of *crawlers* genoemd. Deze speurders grazen het internet af op basis van eerder – door de aanbieder van de zoekmachine – vastgestelde criteria. De aldus gevonden informatie wordt geüniformeerd en gestructureerd, waarmee de basis wordt gelegd voor de vindbaarheid ervan. In een volgende stap wordt de informatie geïndexeerd. Bij deze indexering wordt bepaald wat relevante woorden of combinaties van woorden zijn en wordt niet-relevante informatie, zoals stopwoorden en leestekens, verwijderd. In deze fase wordt ook de informatie dusdanig gestroomlijnd dat bijvoorbeeld verschillen tussen enkelvoud en meervoud, of die ontstaan zijn als gevolg van verbuigingen, een identiek zoekresultaat opleveren. Bepaalde herkenbare woorden, zoals namen van personen en basisbegrippen, worden eventueel geïdentificeerd. De resterende informatie wordt vervolgens nog 'gewogen' door te kijken naar de frequentie van woorden binnen een tekst en de contextuele relevantie of betekenis. Deze verrijkte informatie vormt het uiteindelijke basismateriaal van de zoekmachine.

Lang niet alle informatie die op het internet aanwezig is, wordt door zoekmachines gevonden en geïndexeerd. De schattingen over hoeveel informatie van de aanwezige informatie wel of niet wordt geïndexeerd lopen uiteen. Lage schattingen komen op zo'n twintig procent, hoge schattingen zitten boven de zestig procent. Een en ander doet niet af aan de constatering dat slechts een deel van de aanwezige informatie ook geïndexeerd is of kan worden. Dit heeft verschillende oorzaken. Een deel van de informatie zit verborgen in bestanden die niet gemakkelijk geïndexeerd kunnen worden, zoals tekst in grafische bestanden. Zoekmachines worden echter steeds intelligenter en zijn in staat om bestanden te analyseren en bijvoorbeeld in grafische bestanden tekst en zelfs gezichten te herkennen. Ook bestaat er informatie waarvan de aanbieders niet willen dat deze in zoekmachines wordt genomen. Nieuwsinformatie die snel wordt ververst, leent zich bijvoorbeeld niet voor opname in zoekmachines omdat de informatie alweer verouderd is (er gaan soms maanden voorbij voordat een spider op zoek naar informatie weer langskomt). Verder is er informatie die via het internet wel toegankelijk is, maar zelf niet op het internet aanwezig is, zoals informatie die is opgeslagen in databases. Bovendien groeit het internet nog steeds gestaag.

FIGUUR 5.2

### DE WERKING VAN EEN ZOEKMACHINE



Bij het raadplegen van een zoekmachine wordt grotendeels een proces gebruikt dat omgekeerd is aan het indexeringsproces. De eindgebruiker formuleert een zoekvraag die door de zoekmachine wordt ontleed en geanalyseerd. In dit proces worden weer niet-relevante elementen (zoals stopwoorden) verwijderd, wordt gekeken naar onderlinge relaties tussen zoektermen (die al in de zoekvraag kunnen zijn aangegeven in de vorm van Booleaanse vergelijkingen (AND, OR, NOT)) en het relatieve belang van de ingevoerde zoektermen wordt in kaart gebracht. Dit resulteert dan in een aantal zoekresultaten, die worden afgebeeld op het scherm van de gebruiker (figuur 5.2).

Dit model van verzamelen, ordenen en beschikbaar stellen van informatie is slechts een reflectie van de werkelijkheid. Wat er echt gebeurt voordat een zoekresultaat beschikbaar komt, is zeer complex en wordt in belangrijke mate gekenmerkt door de vele subjectieve elementen die in het proces verweven zijn.

### Manipulatie door zoekmachines

Manipulatie van zoekresultaten doet zich ten minste in twee vormen voor. Er is manipulatie door de zoekmachine maar ook door de aanbieder van informatie die uit is op een zo hoog mogelijke ranking in de lijst met zoekresultaten.

Allereerst de aanbieder van de zoekmachine. Hij is degene die de criteria vaststelt op basis waarvan de op het internet aanwezige informatie wordt verzameld, geordend en beschikbaar gesteld. Informatie die niet wordt gezocht, wordt niet gevonden. Als de zoekrobot die het internet afgraast is geïnstrueerd om bepaalde informatie te negeren, dan zal deze informatie nimmer als het resultaat van een zoekactie kunnen verschijnen. De analyse van een zoekvraag en het te geven antwoord worden bepaald door de instructies die door de aanbieder van de zoekmachine zijn meegegeven. Deze instructies zijn het ware geheim van de smid en het ultieme manipulatie instrument. Men zou hier een vergelijking kunnen maken met het geheime recept van Coca-Cola.

Er zijn zoekmachines die de mogelijkheid bieden om een hoge plaats op de lijst van zoekresultaten te 'kopen'. Hiervan bestaan verschillende varianten. De meest simpele is die waarbij de plek letterlijk wordt gekocht. Andere zoekmachines bieden bijvoorbeeld de mogelijkheid dat pagina's van betalende partijen met voorrang worden geïndexeerd en daardoor hoger eindigen op de lijst van zoekresultaten.

Sommige zoekmachines geven, al dan niet met gebruikmaking van speciale filters, om politieke of commerciële redenen bepaalde resultaten niet weer. Zo zouden zoekmachines bepaalde nieuwsinformatie niet beschikbaar stellen bij zoekvragen die afkomstig zijn uit specifieke landen. Maar het is ook mogelijk dat er juridische belemmeringen zijn om zoekresultaten ter beschikking te stellen. Landen als Frankrijk en Duitsland hebben bijvoorbeeld strikte regels met betrekking tot de promotie of distributie van Nazi-gerelateerd materiaal. Europese regels richten zich meer en meer op tegengaan van de promotie van internetcriminaliteit en terrorisme, hetgeen eveneens een effect kan hebben op de geoorlooftheid van zoekresultaten. Verder zijn er regelmatig gerechtelijke beslissingen over auteursrecht, handelsnaamrecht en oneerlijke handelspraktijken die gevolgen hebben voor de werkwijze van zoekmachines. Ten slotte is er nog het fenomeen dat zoekresultaten kunnen verschillen afhankelijk van de locatie van de vrager of de gebruikte versie van de zoekmachine (zoekvragen op google.nl, google.fr of een andere Google-landenversie leveren geen identieke resultaten op).

Een ander punt waarover nog steeds veel onduidelijkheid bestaat is de 'human factor' in de totstandkoming van het zoekresultaat. Vanzelfsprekend is het opstellen van de instructies voor

de zoekrobot, de indexering, de analyse van de zoekvraag, et cetera gebaseerd op gewoon mensenwerk, maar ook de uitkomsten van het proces zouden worden geanalyseerd en worden bijgesteld. Verder zou er onder meer gewerkt worden met panels om aan verwachtingspatronen te kunnen beantwoorden en zouden mogelijke zoekresultaten handmatig worden aangepast door eigen medewerkers op basis van nader geformuleerde al dan niet subjectieve criteria.

## Manipulatie door informatieaanbieders

Een tweede bron van manipulatie zijn de aanbieders van informatie. Zij kunnen dat doen door in sommige gevallen voor een hoge plaatsing te betalen of door rechtstreeks invloed uit te oefenen op de aanbieder van de zoekmachine, maar vaker wordt door het kundig inrichten van de eigen webinformatie een dusdanig profiel gecreëerd dat de informatie door de zoekmachines hoog op de lijst van zoekresultaten wordt vermeld. Daarbij wordt getracht om te anticiperen op de instructies van de zoekrobot en het indexeringsproces door de zoekmachine (voorzover dat bekend is). Een klassiek voorbeeld is het manipuleren van de eigen *metatags* (de woorden die aangeven wat de belangrijkste kenmerken zijn van de te verstrekken informatie) door het toevoegen van aantrekkelijke zoekwoorden die op zich niets met de eigen dienstverlening te maken hebben (zoals voetbal, porno of de merknamen van concurrenten). Zoekmachines worden echter steeds slimmer en zijn in de regel in staat om de effecten van gemanipuleerde *metatagging* te 'neutraliseren'. Inmiddels worden daarom meer geavanceerde methodes ingezet om de aandacht te vergroten. Zo worden nep-sites opgezet die veel verwijzingen bevatten naar de eigen site om aldus de plaats op ranglijsten te beïnvloeden. Of populaire sites worden gekopieerd en onzichtbaar opgenomen in de eigen site, zodat de nietsvermoedende gebruiker in plaats van bij de door hem beoogde site op een andere plek terechtkomt. Deze en andere vormen van beïnvloeding c.q. misleiding staan bekend onder begrippen als *spamdexing*, *cloacking*, *linkfarming*, *webring*, *redirects*, *doorway pages* en *pagejacking*. In alle gevallen gaat het om methodes om de plaatsing in de zoekresultaten te verbeteren.

De verschillende beïnvloedingstechnieken worden weliswaar bestreden door de zoekmachines, maar lang niet altijd succesvol. Bij Google is de ultieme sanctie uitsluiting van de overtreder, wiens pagina's dan niet meer worden geïndexeerd. De betreffende partij is dan via de zoekmachine onvindbaar geworden. De overtreders zijn niet alleen louche partijen, maar ook de overheid en gerenommeerde bedrijven die gebruikmaken van externe bureaus om de zoekresultaten te optimaliseren. Rond dit optimaliseren van zoekresultaten is een hele industrie ontstaan. Onder de noemer *search engine marketing* (zoekmachinemarketing) worden door bedrijven diensten aangeboden die er op zijn gericht om een hogere plaatsing te krijgen. Zij worden ook wel eufemistisch SEO's genoemd: *search engine optimizers*. Aanbieders van zoekmachines hebben in de regel voorwaarden met betrekking tot SEO-praktijken, die er op neerkomen dat bepaalde vormen van manipulatie worden toegestaan, danwel niet actief worden bestreden.

## Data-honger

De functionaliteit van zoekmachines wordt in hoge mate beïnvloed door de aard en omvang van de onderliggende data. Zoekmachines vergaren niet alleen informatie over wat op het internet aanwezig is, maar koppelen in voorkomende gevallen deze informatie aan de gegevens die zij uit andere bron hebben over degenen die zoekopdrachten plaatsen.

Naast de informatie die op het internet aanwezig is, wordt er aanvullende informatie van gebruikers gegenereerd of toegevoegd. Deze bestaat uit informatie die door gebruikers over henzelf ter beschikking is gesteld. Dat is allereerst de zoekopdracht. Een zoekopdracht bevat in begin-

sel zowel informatie over de gebruiker (zijn persoonlijke voorkeur al dan niet gecombineerd met persoonlijke gegevens) als over de informatie die hij zoekt. Op grond hiervan is een eerste profiel te maken van de gebruiker. Meer zoekopdrachten en bijbehorende resultaten leveren een steeds gedetailleerder beeld op.

Daarnaast wordt informatie over gebruikers verkregen doordat deze bijvoorbeeld formulieren invullen waarbij persoonlijke voorkeuren worden opgegeven maar waarbij ook toestemming wordt gegeven tot het gebruik van deze gegevens en andere gegevens die door gebruikers worden gegenereerd. Gebruikers kunnen onder meer toestemming geven tot het analyseren van persoonlijke bestanden zoals e-mails (dit is het geval bij GMail, de e-maildienst van Google) of data die beschikbaar komen bij het gebruik van (online) applicaties (Google Desktop, Picasa, Google Docs & Spreadsheets).

Indien alle genoemde gegevens worden vastgelegd ontstaat een gigantische database, waarvan de omvang onder meer wordt bepaald door factoren zoals:

- a) het tijdstip waarop begonnen is met het vastleggen van de data
- b) de selectie van de data
- c) de duur van de vastlegging
- d) de herwaardering van de data
- e) de vernietiging van geaggregeerde data

Ofschoon dit fenomeen als zodanig niet onbekend is - datawarehousing en datamining zijn gekende begrippen - is over de data die door zoekmachines worden vastgelegd relatief weinig bekend. Zoekmachines laten zich maar zeer beperkt uit over dit onderdeel van hun activiteiten. Dat ligt gezien de gevoeligheid ervan ook wel voor de hand.

### Verticale integratie

Zoekmachines streven in voorkomende gevallen naar verticale integratie. Dit houdt in dat de zoekmachines zich bezighouden met met content samenhangende exploitatierechten en/of het in eigendom of anderszins verkrijgen en controleren van content. Ook in dit verband is Google een mooi voorbeeld. Het bedrijf bouwt een database van de wereldliteratuur op door het digitaliseren van de inhoud van bibliotheken (Google Books). Werken waarop geen rechten berusten worden integraal toegankelijk gemaakt. Van boeken die nog auteursrechtelijke bescherming genieten wordt alleen een beperkt deel ('snippet') zichtbaar. Een tweede voorbeeld is de acquisitie van YouTube. Op deze site worden door bedrijven en particulieren video's achtergelaten die door internetgebruikers bekeken kunnen worden. Een derde voorbeeld zijn de activiteiten van Google op het gebied van geo-informatie (Google Earth/Google Maps). Vergelijkbare ontwikkelingen zijn zichtbaar bij andere grote zoekmachines.

Naast de verwerving van content, gaan zoekmachines bijzondere c.q. preferente relaties aan met informatieleveranciers en verschaffen zich daarbij toegang tot de gebruikersdata bij deze informatieleveranciers. De betreffende partijen zien de voordelen terug in de vorm van zoekresultaten en daarmee grotere aandacht voor hun diensten. Onder meer via *revenue sharing*-modellen wordt ook de zoekmachine-aanbieder er eventueel beter van.

## De zoekmachinemarkt

Er zijn vele soorten zoekmachines. Sommige richten zich op een bepaalde gespecialiseerde markt (www.findsounds.com of scholar.google.nl) of fungeren als metazoekmachine door resultaten afkomstig van verschillende zoekmachines te combineren (bijvoorbeeld vinden.nl). Hier gaat de aandacht in de eerste plaats uit naar zoekmachines die zich op het algemene publiek richten, zoals Google, Yahoo of Ilse.

Nog niet zolang geleden, aan het begin van deze eeuw, waren er al zeer veel zoekmachines actief en was de algemene veronderstelling dat concurrentie tussen zoekmachines de markt zou disciplineren. Zowel informatie-aanbieders als gebruikers zouden hierbij baat hebben. Ofschoon het aantal zoekmachines nog steeds erg groot is, heeft er een enorme verschuiving plaats gevonden wat betreft marktaandeel.

Recente informatie over de Amerikaanse markt toont dat Google, Yahoo, Live Search (MSN) en Ask samen een marktaandeel hebben van ruim 98 procent (tabel 5.2). Alle andere zoekmachines moeten het doen met de resterende 1,7 procent. Google is de onbetwiste marktleider met een aandeel van ruim eenderde.

TABEL 5.2

### AANDEEL ZOEKOPDRACHTEN IN DE VERENIGDE STATEN VOOR DE GROOTSTE ZOEKMACHINES

(IN PROCENTEN)

ZOEKMACHINE	2006	2007	2008
Google	58.3	64.1	67.3
Yahoo	22.3	21.3	20.3
Live Search (MSN)	13.1	9.0	6.7
Ask	4.0	3.5	4.1

Bron data: Hitwise

Er is een interessant verschil tussen de Verenigde Staten en Europa. Ondanks dat Google een Amerikaanse onderneming is, is de dominantie in Europa nog groter dan in de Verenigde Staten. Data uit 2007 over de markt in Duitsland, Frankrijk en het Verenigd Koninkrijk (VK) spreken wat dat betreft boekdelen:

TABEL 5.3

### AANDEEL ZOEKOPDRACHTEN IN DRIE EUROPESE LANDEN VOOR DE GROOTSTE ZOEKMACHINES

(IN PROCENTEN)

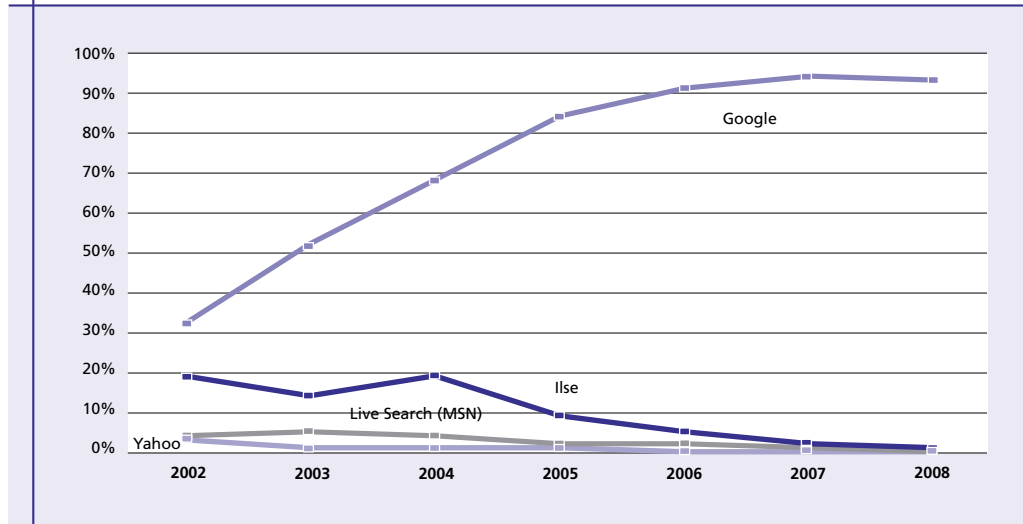
ZOEKMACHINE	DUITSLAND	FRANKRIJK	VK
Google	88,5	88,8	79,4
Yahoo	3,4	3,1	7,7
Live Search (MSN)	1,4	2,5	5,3

Bron data: Webhits / Xitimonitor / Hitwise

In Duitsland, Frankrijk en het Verenigd Koninkrijk zien we een overeenkomstig beeld. Google heeft een marktaandeel van 80 tot 90 procent en de drie grote spelers in de Verenigde Staten hebben gezamenlijk zo'n 95 procent van de markt in handen.

FIGUUR 5.3

**AANDEEL ZOEKOPDRACHTEN IN NEDERLAND BIJ DE VIER GROOTSTE ZOEKMACHINES**



Bron data: Checkit

In Nederland is de situatie nog wat extremer (figuur 5.3). Google staat bovenaan met een aandeel van ruim 90 procent. Yahoo en Live Search die elders in Europa nog wat procentjes pakken hebben in Nederland geen in procenten meetbaar resultaat meer. Ilse dreigt zich hierbij te moeten gaan voegen.

Er zijn hier en daar nog wat landen in Europa waar de geschetste verhoudingen niet opgaan en een lokale zoekmachine nog een belangrijk marktaandeel heeft, maar dat soort situaties komt steeds minder vaak voor.

**Bedrijfsmodel van zoekmachines**

De inkomsten van zoekmachines komen voornamelijk uit één bron: reclame. Nemen we Google als voorbeeld, dan blijkt dat deze zoekmachine in 2007 16,4 miljard dollar aan advertentie-inkomsten had. Slechts 181 miljoen dollar (oftewel een magere 1 procent) komt uit andere bronnen. De reclame-inkomsten zijn vrijwel uitsluitend internet-gerelateerd. De verhoudingen zijn sinds de beursgang van Google in 2004 niet veel anders geweest.

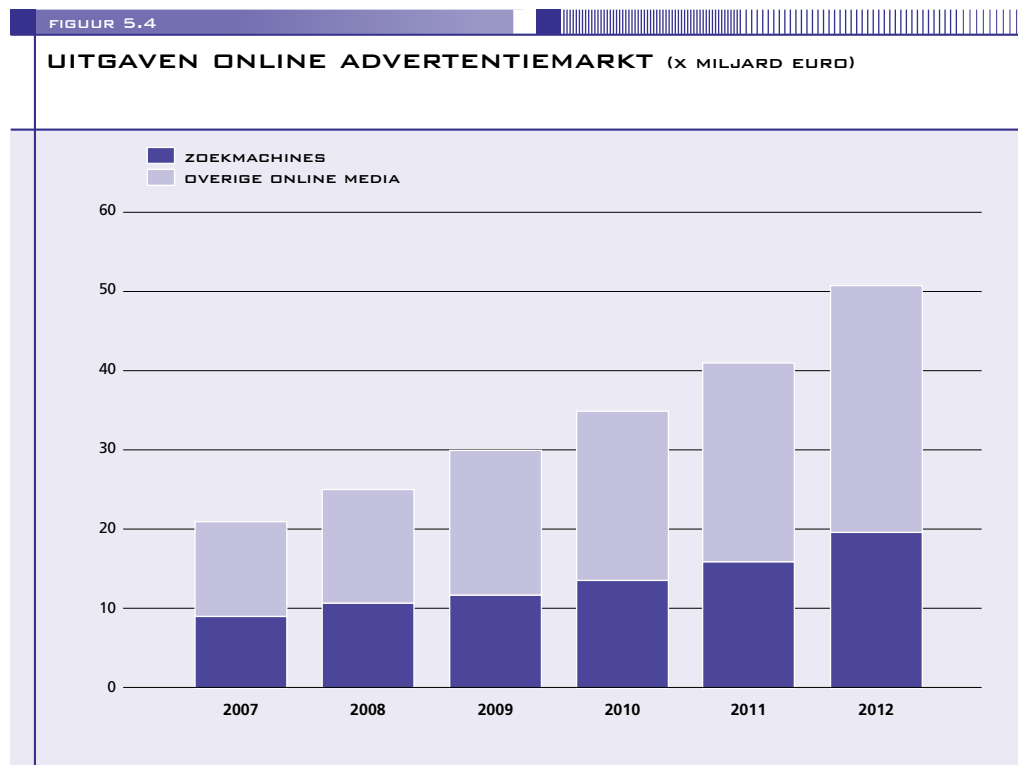
De reclame-inkomsten worden in de eerste plaats gegenereerd door middel van 'Google Adwords'. Via Adwords kunnen adverteerders zelf hun advertenties samenstellen en aangeven hoeveel geld zij er aan willen spenderen. Vervolgens worden ze afgerekend aan de hand van het aantal keren dat de advertentie wordt aangeklikt. De advertenties verschijnen op de website van Google naast de resultaten van een zoekopdracht. Google bepaalt welke advertentie wanneer verschijnt en doet dat primair in relatie tot de zoekopdracht.

De tweede bron van inkomsten bestaat uit het plaatsen van aangeboden advertenties op websites van derden. Dit gebeurt via het programma 'AdSense', dat twee varianten kent: 'AdSense for search' en 'AdSense for content'. Bij 'AdSense for search' worden advertenties geplaatst in relatie tot zoekopdrachten op de websites van derden. Bij 'AdSense voor content' worden advertenties gerelateerd aan de inhoud van websites. Ten aanzien van AdSense kent Google



een *revenue sharing*-model, waarbij een deel van de gegenereerde advertentie-inkomsten naar de aanbieders van de informatie gaat. Deze zijn dus in de positie om bij de samenstelling van de inhoud van hun website hiermee rekening te houden en de inhoud te 'optimaliseren'.

Het gaat bij de markt van online-reclame om een enorme groeimarkt, waarvan de zoekmachines hun deel willen binnenhalen. In de Verenigde Staten vertegenwoordigt de markt van internet-reclame in 2007 al 7,4 procent van de totale advertentiemarkt, goed voor 21,1 miljard dollar. De verwachting is dat dit marktaandeel in 2012 al meer dan verdubbeld zal zijn.



De online advertentiemarkt bestaat niet alleen uit inkomsten die gerelateerd zijn aan zoekresultaten, maar ook uit andere reclamevormen zoals advertenties, banners en reclamefilmpjes (figuur 5.4). Wat betreft de inkomsten uit 'search'-advertenties: het overgrote deel daarvan gaat volgens de schattingen van eMarketer naar Google (2007: 75 procent; 2006: 60 procent).

De omvang van de online advertentiemarkt is een van de redenen waarom Google onlangs het bedrijf DoubleClick overnam. DoubleClick is een grote partij in de markt van internet advertenties en banners en vult de bestaande advertentie-activiteiten van Google goed aan. Samen zouden zij goed zijn voor 69 procent van de online advertentiemarkt.

De online advertentiebestedingen in Europa en Nederland zijn aardig in lijn met de Amerikaanse ontwikkelingen. Volgens IAB Europe komen de Europese bestedingen over 2007 uit op 11,2 miljard euro, wat een sterke groei betekent ten opzichte van 2006 toen het nog om een bedrag van 8 miljard euro ging. Daarmee komt de Europese markt qua omvang langzaam maar zeker in de buurt van de Amerikaanse (mede geholpen door de koersverhouding euro/dollar). De netto mediabestedingen van 'online display advertising' voor Nederland zijn in 2007 uitgekomen op 26 procent van de totale kosten aan internet. De netto mediabestedingen van 'zoekmachinemarketing' nemen in datzelfde jaar 50 procent in beslag.

Waarschijnlijk is er nog het nodige af te dingen op de genoemde cijfers en zijn ze niet altijd vergelijkbaar, er zijn wel duidelijke trends waarneembaar: de markt van online advertentiebestedingen maakt een snelle groei door en neemt een steeds groter deel van de totale advertentiemarkt in beslag. Zoekmachines worden primair uit deze online advertentiemarkt gefinancierd.

Dat het moeilijk is om andere inkomsten dan (online) advertenties te vinden, geeft de eerdere onderverdeling in inkomstenstromen bij Google aan. Al sinds 2004 bedragen de inkomsten uit 'Licensing & other revenue' niet meer dan 1 procent. In absolute termen zijn ze weliswaar gestegen van bijna 46 miljoen dollar (2004) naar 112 miljoen dollar (2007), maar het blijft een druppel op een gloeiende plaat.

Google is sowieso een mooi voorbeeld van hoe men – naast overnames in de eigen core-sector zoals die van DoubleClick - het business model verder wil uitbreiden. De wijze waarop het bedrijf advertenties uitvent op het internet probeert men toe te passen in andere sectoren. Zo wordt het Adwords-systeem inmiddels ook toegepast om reclamezendtijd te verkopen aan radio-adverteerders (zogenaamde 'AudioAds'). Al meer dan 1600 radio stations – waaronder de 675 stations van de keten Clear Channel – zouden van deze dienst gebruik maken.

### **Plaats van de zoekmachine in de mediasector**

Waar bevinden zich de activiteiten van de zoekmachine vanuit het perspectief van de mediasector? Op deze vraag is geen eenvoudig antwoord mogelijk en het hangt bijvoorbeeld af van de vraag of gekozen wordt voor een smalle of een brede benadering. In een smalle benadering gaat het om de zoekmachine als middel om op het internet aanwezige informatie te ontsluiten: er is aanbod van informatie en er is vraag naar informatie. Zoekmachines brengen vraag en aanbod met elkaar in contact. De over de gebruiker verzamelde informatie wordt gebruikt om zijn zoekvraag steeds beter te analyseren teneinde een zo optimaal mogelijk zoekresultaat te bewerkstelligen. Bij een bredere benadering zou de positie van de aanbieder c.q. het onderliggende businessmodel als uitgangspunt kunnen worden genomen. We zien dan dat de zoekmachine in de eerste plaats een middel is om 'eyeballs' te verkopen aan adverteerders en informatie-aanbieders, waarbij de over de gebruiker vergaarde informatie er primair toe dient om het inkomstenmodel te optimaliseren.

De zoekmachinemarkt wordt beheerst door partijen met een bredere benadering. Dit impliceert dat het commerciële oogmerk voorop staat waarbij adverteerders en informatie-aanbieders uitgangspunt zijn. Het gaat er niet, dan wel niet uitsluitend om dat het zoekresultaat gezien van uit de gebruiker gemaximaliseerd wordt. Immers niet de gebruiker betaalt, maar de adverteerder of informatie-aanbieder. Het is begrijpelijk dat er wel vergelijkingen worden gemaakt tussen het omroep-model en de wijze waarop de markt van zoekmachines functioneert. De analogie gaat niet in alle opzichten op, maar er zijn voldoende elementen die vergelijkbaar zijn zoals de noodzaak tot gebruikersaanblik en de afhankelijkheid van adverteerders.

## Zoekmachines in recht en beleid

Er zijn veel redenen waarom zoekmachines voldoende aandacht moeten krijgen in recht en beleid. Zoekmachines blijken een centrale rol te vervullen bij het ontsluiten van informatie op het internet. Tegelijkertijd is er sprake van manipulatie van zoekresultaten, van het vergaren en verwerking van enorme hoeveelheden persoonlijke gegevens en van afhankelijkheid van commerciële overwegingen bij de exploitatie van zoekmachines. Daar komt nog bij dat wat betreft de Europese en Nederlandse markt er sprake is van één dominante partij, de top-3 in de regel meer dan 95 procent marktaandeel heeft, waarbij bovendien deze top-3 uitsluitend bestaat uit niet-Europese en niet-nationale aanbieders.

Vanuit het recht vertaalt een en ander zich naar fundamentele vragen ten aanzien van de informatievrijheid en de privacy. De positie van zoekmachines is dat zij enerzijds invloed hebben op de informatie die – de facto – beschikbaar is en anderzijds beïnvloeden of de informatie ook daadwerkelijk wordt ontsloten voor gebruikers. Het proces van informatievoorziening via het internet komt aldus in belangrijke mate buiten de controlesfeer van zowel de informatie-aanbieder als de informatievrager te staan in zowel kwantitatieve als kwalitatieve zin. De kritiek dat informatie op het internet ook beschikbaar is zonder de tussenkomst van zoekmachines is op zich niet onjuist, maar staat op gespannen voet met de realiteit van de dag, die een hoge afhankelijkheid van zoekmachines kent. De informatievrijheid is in het geding, maar krijgt vooralsnog weinig aandacht in de beleidsvorming. Wel is er een trendbreuk te constateren met betrekking tot de privacy. Langzaam maar zeker leeft de belangstelling voor privacy weer op. Dat heeft niet alleen te maken met het feit dat de vraag naar en het gebruik van privacygevoelige informatie een grote vlucht heeft genomen (zoals bij de bestrijding van criminaliteit en terrorisme). Privacy speelt meer dan gemiddeld bij zoekmachines zoals is geïllustreerd bij de beschrijving van het businessmodel, dat grotendeels leunt op het verschaffen van zo nauwkeurig mogelijke persoonlijke informatie voor commerciële doeleinden.

Naast informatievrijheid en privacy, vragen de marktverhoudingen aandacht. De structuur van de markt indiceert dat toetreding van nieuwe partijen geen eenvoudige zaak is en risico's verbonden aan marktmacht (kunnen) optreden. Marktfragen worden niet alleen ingegeven door de bestaande posities van partijen, maar eveneens door de eerder beschreven processen van verticale en/of horizontale integratie en schaalvergroting.

Met de beleidsmatige en juridische inkadering van zoekmachines moet in Europa (en dus ook in Nederland) nog een begin worden gemaakt. Activiteiten van zoekmachines vallen grotendeels buiten de bestaande regulering van de media- of telecommunicatiesector. Noch de Mediawet noch de Telecommunicatiewet ziet toe op zoekmachines. Hetzelfde geldt voor de Europese regels voor de communicatiesector en audiovisuele mediadiensten. Algemene regels inzake commerciële activiteiten (zoals oneerlijke handelspraktijken), privacy en mededinging zijn natuurlijk wel van toepassing, maar de vraag is of dat de juiste en afdoende instrumenten zijn. Veel van deze algemene regels hebben namelijk een ex-post karakter, terwijl er goede redenen kunnen zijn om vooraf al waarborgen in te bouwen. Maar ook is de vraag of generieke regelgeving zich wel zo gemakkelijk laat toepassen. Binnen het mededingingsrecht zijn concepten als privacy en vrijheid van meningsuiting moeilijk inpasbaar, zo wijst de praktijk van het mededingingstoezicht uit.

Een en ander neemt niet weg dat de sectorspecifieke regels zoals die gelden voor de media- en telecommunicatiesector een inspiratiebron kunnen zijn voor mogelijke regulatoire inkadering van zoekmachines. Toegang tot zoekmachines voor aanbieders van informatiediensten zou bijvoorbeeld onderhevig kunnen zijn aan criteria als een verbod op uitsluiting en een recht op toegang tegen non-discriminatoire en objectieve voorwaarden. Transparantievoorwaarden zouden zich kunnen richten op het inzichtelijk maken van bedrijfsprocessen die bepalen op

welke wijze zoekresultaten totstandkomen. Onafhankelijke derden zouden deze processen kunnen controleren. Aanbieders van zoekmachines zouden aan bepaalde kwaliteitseisen kunnen voldoen die al dan niet via co-regulering (bijvoorbeeld een keurmerk) de gebruiker meer zekerheid bieden. Voor wat betreft de verantwoordelijkheid van zoekmachines (in welke mate is er mogelijk sprake van redactionele verantwoordelijkheid) kan geput worden uit het raamwerk dat bestaat met betrekking tot de verantwoordelijkheden rond het verzorgen van audiovisuele mediadiensten. Hetzelfde kader zou aanknopingspunten kunnen bieden voor de wijze waarop tegen de commerciële exploitatie (advertenties, banners) moet worden aangekeken. Een bestaande privacyregulering zou als oriëntatie kunnen dienen voor eventuele interventie bij het gebruik en de opslag van persoonsgegevens. De wijze waarop met data wordt omgegaan zou zich eveneens lenen voor toetsing door onafhankelijke derden.

Twee belangrijke kanttekeningen bij de verdere gedachtevorming over de mogelijke regulering van zoekmachines. In de eerste plaats dienen regels niet zomaar te worden ontleend aan bestaande reguleringskaders. Telkens zal moeten worden nagegaan of de motieven die aan bestaande regels ten grondslag liggen nog steeds valide zijn, waarna vanuit deze onderliggende beginselen reguleringsmodellen voor zoekmachines zouden kunnen worden opgebouwd. In de tweede plaats behoort er voldoende aandacht te zijn voor het feit dat zoekmachines bij uitstek een voorbeeld zijn van 'geconvergeerde dienst' waarbinnen onder meer telecommunicatie-, media- en privacyvraagstukken een rol spelen. Er dient een evenwichtige samenhang te zijn in regelgeving, beleid en toezicht die deze complexiteit reflecteert. Afzonderlijke regel- en toezichthokjes passen daar niet goed bij.

---

<sup>1</sup> Prof. dr. N.A.N.M. van Eijk is als hoogleraar verbonden aan het Instituut voor Informatierecht (IViR, Universiteit van Amsterdam, [www.ivir.nl](http://www.ivir.nl)). Hij wijdde zijn inaugurele rede aan de problematiek van zoekmachines. Deze bijdrage is daarop mede gebaseerd en er deels aan ontleend. Omwille van de leesbaarheid is grotendeels afgezien van het opnemen van voetnoten. Relevante verwijzingen zijn onder meer terug te vinden in de inaugurele rede van 2005 (zie: [http://www.ivir.nl/publicaties/vaneijk/oratie\\_zoekmachines.pdf](http://www.ivir.nl/publicaties/vaneijk/oratie_zoekmachines.pdf)).