

# Computercursus

# Computercursus

Zoeken

Zoeken Les 2

Zoekmachines (1)  
Les 2

Zoekmachines (1)



[www.jongleertoud.nl](http://www.jongleertoud.nl)



## Zoekmachines of search engines

- Wat is een zoekmachine?
  - Reikwijdte
  - Waarin verschillen zoekmachines van elkaar?
  - Reclame
  - Een aantal zoekmachines op een rij
  - Nieuwe ontwikkelingen
- 

### Wat is een zoekmachine?

Een bekende manier om informatie op het World Wide Web op te sporen is gebruik maken van zoekmachines (= search engines).

In tegenstelling tot de onderwerpsguides, die door deskundigen worden samengesteld, worden er bij zoekmachines met behulp van computerprogramma's omvangrijke indexen aangemaakt. In de indexen probeert men elk (belangrijk) woord van elke pagina van het hele WWW onder te brengen. Gezien de enorme hoeveelheid Webpagina's en de dagelijkse aanwas van nieuwe Webpagina's, is dit bij voorbaat een onmogelijkheid.

Met een zoekmachine kun je, door het invullen van een of meer zoekwoorden, zoeken naar webpagina's op het World Wide Web.

Zoekmachines maken gebruik van computerprogramma's, zogenaamde spiders of crawlers, die het hele Web afspeuren op zoek naar webpagina's.

Ze maken een kopie van elke webpagina die ze tegenkomen en slaan die op in een database. Wanneer je gebruik maakt van een zoekmachine, zoek je dus eigenlijk niet direct op het Web, maar in de database met webpagina's die de zoekmachine heeft aangelegd.

De spiders of crawlers speuren het Web af op zoek naar nieuwe webpagina's, maar ook op zoek naar nieuwe informatie in pagina's die ze al eens eerder hebben bezocht.

Tot slot worden de opgeslagen webpagina's geïndexeerd: er wordt van elke webpagina een lijst aangelegd met de (belangrijkste) woorden die voorkomen in de pagina.

Als je nu de zoekmachine vraagt te zoeken naar een bepaald zoekwoord, kijkt de zoekmachine in de index waar dit zoekwoord voorkomt en welke webpagina's er bijhoren. Een enorme klus natuurlijk. Enkele seconden later krijg je de webpagina's gepresenteerd. En dat kunnen er bij een zoekmachine vele duizenden zijn, want de zoekmachine selecteert niet op kwaliteit, maar kijkt alleen in welke webpagina's een bepaald woord voorkomt.

Houd dus goed in de gaten dat:

- je alleen zoekt in het bestand met documenten dat door de zoekmachine is aangelegd en niet op het hele WWW;
- je alleen kunt zoeken naar de documenten met behulp van woorden uit die documenten.

### Reikwijdte

Onderzoek uit 1999 van het NEC Research Institute toont aan dat de elf meest gebruikte zoekmachines samen slechts 42% van het Web bestrijken. Meer dan de helft van het Web is dus onbereikbaar voor zoekmachines.

Meer recente onderzoeken, besproken door Danny Sullivan in zijn artikel *Avoiding The Search Gap*, bevestigen dit beeld.

Zoekmachines blijven echter het enige instrument waarmee je door middel van het opgeven van zoekwoorden informatie op het Web kunt opzoeken.

## Waarin verschillen zoekmachines van elkaar?

Er zijn vele duizenden zoekmachines die allemaal een gedeelte van het Web afzoeken. Sommige lijken op elkaar, andere zien er weer helemaal anders uit. Waarin onderscheiden deze zoekmachines zich nu van elkaar?

- de snelheid van de spider. Hoeveel webpagina's kan de spider bezoeken en verwerken?
- de techniek van de spider. Wordt alleen de homepage van een website bezocht of speurt de spider dieperliggende webpagina's af?
- de omvang van de index (lijst van woorden waarmee de zoekmachine de eigen database van websites afzoekt)
- de omvang van de database waarin kopieën van webpagina's zijn opgeslagen. Welk gedeelte van het Web is met de zoekmachine te doorzoeken?
- de snelheid waarmee de zoekmachine de eigen database afzoekt.
- de gebruiksvriendelijkheid van de zoekmachine. Op welke manieren kan de gebruiker zoeken: eenvoudig, geavanceerd?
- lay-out

Een klein verschil in een van deze elementen zal effect hebben op de resultaten van een zoekactie. Dit is een van de redenen waarom resultaten van zoekmachine tot zoekmachine verschillen wanneer met dezelfde zoekwoorden wordt gezocht. Raadpleeg daarom altijd meer dan een zoekmachine.

Door de ontwikkelaars van zoekmachines wordt uit concurrentieoverwegingen belangrijke informatie over de exacte werking van hun zoekmachine achtergehouden. De aanwijzingen die via de helpschermen worden gegeven zijn niet voldoende om de mogelijkheden van een zoekmachine volledig te benutten. Datgene wat men de gebruiker via helpschermen aanbiedt zijn dan ook slechts zoektips en trucs. Een andere hindernis om details over de werking van zoekmachines te achterhalen wordt gevormd door de vele veranderingen die de zoekmachines zelf ondergaan. De aangeboden documentatie blijft daarbij vaak achter. Er wordt constant aan de zoekmachines gesleuteld om hen nog sneller, nog precieser, nog geavanceerder te maken dan hun naaste concurrenten. Daarbij verliest men de gebruiker nogal eens uit het oog.

## Reclame

Hoe gebruiksvriendelijker, sneller en omvangrijker een zoekmachine is, des te hogere eisen er gesteld moeten worden aan de zoekmachine zelf en aan de programmeurs van de zoekmachine. Dit maakt de ontwikkeling en het onderhoud van een zoekmachine ook duurder. Tot 2000 was het absoluut ondenkbaar dat men zich als onderneming een plekje bij de eerste 10 gepresenteerde resultaten kon kopen. Op dit moment echter kent elke zoekmachine wel een of andere manier waarop adverteerders, in of in de directe nabijheid van de zoekresultaten, hun boodschap kwijt kunnen. Zie ook Danny Sullivans' artikel *Buying Your Way In To Search Engines*.

### **Welke vormen van reclame zijn inmiddels gangbaar?**

#### **Banner Ads**

Elke grotere zoekmachine kent aan de specifieke zoekwoorden gerelateerde grafische of tekstuele reclameboodschappen. Wanneer je bijvoorbeeld op zoek bent naar informatie over persoonlijke leningen, verschijnt er bij de zoekresultaten een reclametekst van Nationale Nederlanden.

**Content Promotion**

Advertenties, veelal geplaatst aan de rechterkant van de zoekresultaten.

**Paid Inclusion**

Er wordt betaald voor het opnemen van een site in de index van een zoekmachine. Er wordt echter geen garantie afgegeven voor een hoge klassering bij een bepaalde zoekvraag.

**Paid Placement**

In tegenstelling tot bij Paid Inclusion wordt bij een bepaald zoekwoord een hoge klassering in de resultaten gegarandeerd. De meest biedende adverteerder eindigt dan het hoogst op de resultatenlijst. De exacte positie varieert. Soms verschijnen zij boven de echte zoekresultaten, soms er onder.

**Paid Submission**

Er wordt betaald om (eventueel) opgenomen te worden in de index van een zoekmachine. Er wordt echter geen garantie afgegeven dat de site in de index wordt opgenomen. De site wordt door een redactie bekeken en mogelijk met voorrang opgenomen in de index.

## Een aantal zoekmachines op een rij

**Google** <http://www.google.nl/>

Google maakt gebruik van een waarderingssysteem dat websites en andere Internetdocumenten beoordeelt op grond van het aantal links van andere sites naar die betreffende documenten. Deze techniek wordt PageRanking genoemd. Hoe meer webpagina's er verwijzen naar een site, hoe populairder deze site en dus: hoe hoger de website staat op de lijst van zoekresultaten. Wanneer bijvoorbeeld gezocht wordt naar General Motors, dan wordt het resultaat van die zoekactie gebaseerd op de populariteit van sites over General Motors. In het geval van een zoekactie naar General Motors, krijg je als resultaat de homepage van de onderneming General Motors als eerste treffer gepresenteerd. PageRank kijkt daarnaast eveneens naar de gelinkte sites en gaat na, hoeveel van hen teruglinken.

Achter elk zoekresultaat volgt de optie Similar pages / Gelijkwaardige pagina's. Als je dit aanklikt, verschijnt er een lijst met verwante URL's. Bij de homepage van de onderneming Agfa bijvoorbeeld, levert het activeren van Similar pages een opsomming van homepages van concurrerende ondernemingen als Polaroid, Canon en andere branchegenoten op.

Ook kent Google een hogere waardering toe wanneer een hooggewaardeerde site, naar een bepaalde site verwijst.

Hoe de wiskundige formule van PageRanking precies in elkaar zit blijft een groot Google-geheim.

Gebruik Search Tips om meer zoekmogelijkheden van Google te ontdekken.

Wanneer bijvoorbeeld Yahoo! verwijst naar een bepaalde site, dan stijgt deze laatste automatisch in de waardering. Google presenteert dus de mainstream, de meest gangbare en derhalve veelal betrouwbare sites over een bepaald onderwerp. Persoonlijke homepages en obscure sites zul je dus nauwelijks aantreffen.

De tabs boven het zoekvenster bieden op een vriendelijke en eenvoudige manier toegang tot Google's andere databases; afbeeldingen, discussies, nieuws en gids. Onder deze laatste optie bevindt zich de door deskundigen samengestelde onderwerpsgids, gebaseerd op de dmoz.org directory (Open Directory Project).

De Advanced Search-optie is gebaseerd op pull-down menu's, waarbij bijvoorbeeld zoekwoorden op eenvoudige wijze gecombineerd kunnen worden. Op dezelfde pagina zijn links aangebracht naar Google's gespecialiseerde zoekmachines. Ook voorziet Google in een handig woordenboek. Boven de zoekresultaten bevindt zich namelijk een statusbalk, waarin het zoekwoord onderstreept en doorklikbaar is gemaakt. Met behulp van Dictionary.com wordt de betekenis van een (zoek)woord aangegeven. Eveneens kan doorgeklikt worden naar een passende thesaurusterm.

Met Google kun je niet alleen zoeken naar HTML-documenten, maar ook naar PDF-, DOC-, TXT- en andere tekstdocumenten. Vooral de PDF-documenten zijn interessant wanneer je op zoek bent naar serieuze documenten. Met name wetenschappelijke publikaties verschijnen vaak in PDF-format. Op deze wijze wordt een belangrijk gedeelte van het Invisible Web zichtbaar gemaakt. Om PDF-bestanden te kunnen lezen moet je eerst Acrobat Reader van Adobe (gratis) downloaden.

### Oefening:



Zoek in Google <http://www.google.com/> naar nederlandstalige PDF-documenten over Oorlog en vrede van Leo Tolstoj/Tolstoy/Tolstoi. (Gebruik de optie Geavanceerd zoeken / Advanced search).

**Yahoo!** <http://www.yahoo.com/>

Yahoo! houdt van een regelmatige metamorfose. Jarenlang was Yahoo! een bekende en veel gebruikte onderwerpsgids met een goed uitgewerkte indeling in onderwerpen. Later gaf Yahoo! de gebruiker ook de mogelijkheid om van een zoekmachine gebruik te maken, wat mogelijk was door het inkopen van zoekresultaten bij andere zoekmachines. Zo kon je Yahoo! niet alleen als onderwerpsgids gebruiken, maar ook als zoekmachine. Nog een stap verder ging Yahoo! door het opkopen van de zoekmachines AltaVista, AllTheWeb en Inktomi. Nu timmert Yahoo! aan de weg als grote zoekmachine, waar je ook gebruik kunt maken van een onderwerpsgids. Maar de nadruk ligt nu toch duidelijk op de zoekmachine.

**Teoma** <http://www.teoma.com/>

Is ontwikkeld door Rutgers University, maakt evenals Google gebruik van zgn. link-populariteit.

De zoekresultaten worden op een drietal wijzen gepresenteerd. Onder het kopje Results: worden de gangbare, direct aan het zoekwoord gekoppelde resultaten gepresenteerd. Refine: geeft suggesties om de zoekactie te verfijnen. Bij Resources vind je Webpagina's over verwante onderwerpen.

**Scirus** <http://www.scirus.com/>

Scirus richt zich met name op openbare wetenschappelijke informatie en specifieke wetenschappelijke databases.

**Hotbot** <http://hotbot.com/>

Hotbot is een zoekmachine die de mogelijkheid biedt de afzonderlijke databases van HotBot, Google en Ask Jeeves te doorzoeken. De resultaten worden per zoekmachine gepresenteerd.

Hotbot, ontwikkeld door medewerkers van het tijdschrift Wired Magazine, kenmerkt zich door zijn uitgebreide Advanced zoekopties. In tegenstelling tot de meeste andere zoekmachines kunnen operatoren bijvoorbeeld ook in combinatie met vormopties (images, video, audio etc.) gebruikt worden.

Voor een complexe, duidelijk toegespitste vraag is Hotbot met zijn filters en andere opties om een zoekactie verder te specificeren, het juiste instrument.



### Oefening:

Zoek in de zoekmachine Hotbot <http://hotbot.lycos.com/> naar sites over bibliotheek Universiteit van Tilburg.

**Ask Jeeves** <http://www.ask.com>

Ask Jeeves was in de beginjaren een van de weinige zoekmachines waar je in gewone, natuurlijke taal je vraag kon stellen. Nu ligt bij Ask Jeeves minder de nadruk op Natural Language Processing maar vooral op de door Teoma ontwikkelde zoektechnologie. Ask Jeeves timmert aan de weg met enkele populaire zoekmogelijkheden zoals: Map search, Movie search, White pages search, Dictionary search, Weather search, Product search, News search, Stock search.

**Alltheweb** <http://www.alltheweb.com/>

AllTheWeb heeft jarenlang als zelfstandige zoekmachine gefungeerd, maar is nu opgekocht door Yahoo! AllTheWeb heeft nog wel een eigen interface maar maakt gebruik van de Yahoo!-index. Onduidelijk is hoe lang AllTheWeb nog met een eigen interface zal blijven bestaan voordat het helemaal door Yahoo! wordt opgeslokt.

**Alta Vista** <http://www.altavista.com/>

In de beginjaren van het World Wide Web was de zoekmachine AltaVista de populairste en een van de meest gebruikte zoekmachines. AltaVista is opgekocht door Yahoo! en bestaat niet meer als zelfstandige zoekmachine nu gebruik wordt gemaakt van de grote Yahoo!-index. Er is nog wel een eigen AltaVista-interface maar dat is niet meer dan een mooie verpakking van de Yahoo!- database.



**Oefening:**

Zoek in de zoekmachine Alta Vista (<http://www.altavista.com/>) naar sites over Julia Kristeva. (Vul in Julia Kristeva of Kristeva).

<b>WiseNut</b>	<a href="http://www.wisenut.com/">http://www.wisenut.com/</a>
<b>illumiRate</b>	<a href="http://www.illumirate.com/">http://www.illumirate.com/</a>
<b>Simpli.com</b>	<a href="http://www.simpli.com/">http://www.simpli.com/</a>
<b>Wotbox</b>	<a href="http://www.wotbox.com/">http://www.wotbox.com/</a>
<b>Mooter</b>	<a href="http://www.mooter.com/moot">http://www.mooter.com/moot</a>
<b>Brainboost</b>	<a href="http://brainboost.com/">http://brainboost.com/</a>
<b>NetNose</b>	<a href="http://www.netnose.com/">http://www.netnose.com/</a>
<b>AOL.Search</b>	<a href="http://search.aol.com">http://search.aol.com</a>
<b>Gigablast</b>	<a href="http://gigablast.com">http://gigablast.com</a>

## Nieuwe ontwikkelingen: portals en vortals

Aanvankelijk waren de zoekmachines vooral ontworpen om Internetgebruikers wegwijs te maken in de doolhof van het Web. De afgelopen jaren hebben sommige van deze zoekinstrumenten zich ontwikkeld tot toegangspoorten of portals. Hier bieden zij niet alleen de duizenden gebruikers een toegang tot het Web, maar voorzien hun tevens van gratis e-mail, gepersonaliseerde informatie, onderwerpsgids, maar laten hen ook kennis maken met online winkelen, nieuws en andere services.

Ondernemingen adverteren op zo'n site omdat ze zo een groot aantal mensen weten te bereiken. Om deze reden is de belangstelling voor bedrijven als Yahoo, America Online en Excite toegenomen.

De ervaring wijst uit dat het beheren van een eigen toegangspoort de beste manier is voor Internet-aanbieders om nieuwe klanten te werven.

Meer en meer portals gaan zich toeleggen op een speciaal onderwerp, een aandachtsgebied. Dit is een ontwikkeling die eveneens te bespeuren valt bij zoekmachines, onderwerpsgidsen en metasearchers. Deze zogenaamde vortals of vertical portals bieden toegang tot informatie die is toegespitst op een specifieke marktsector.

Voorbeelden van Portals:

<b>Excite</b>	<a href="http://www.excite.com/">http://www.excite.com/</a>
<b>Lycos</b>	<a href="http://www.lycos.com/">http://www.lycos.com/</a>
<b>About.com</b>	<a href="http://home.about.com/">http://home.about.com/</a>
<b>Microsoft's MSN.com</b>	<a href="http://www.msn.com/">http://www.msn.com/</a>
<b>NBCi</b>	<a href="http://nbc.msnbc.com/nbc.asp">http://nbc.msnbc.com/nbc.asp</a>