

Zoeken op het Web

Inleiding

Om bepaalde documenten op het web te vinden kan je heel wat moeilijkheden ondervinden, soms bekom je zelfs niet het gewenste document. Een van de belangrijkste redenen hiervoor is dat het WWW enorm groot is. Er wordt geschat dat het web momenteel 3 biljoen documenten bevat. Een andere reden hiervoor is dat het web niet geïndexeerd is volgens een bepaalde vocabulary. Een bibliotheek kan bijvoorbeeld op titel zoeken, maar bij het zoeken op Internet naar een bepaalde pagina is het altijd raden welke termen er door iemand zijn gebruikt om zijn web pagina te organiseren of over welk onderwerp het precies gaat.

Als je op het web zoekt naar sites over een bepaald onderwerp, dan ben je niet direct op het web zelf aan het zoeken. Het is niet mogelijk om direct op het web zelf te zoeken. Het web bestaat uit een heel groot aantal webpagina's (die op servers staan), maar het is niet mogelijk dat je computer deze allemaal vindt, of er naar toe gaat.

Search-tools

Als je via een search-tool zoekt, ben je eigenlijk aan het zoeken in de database van de search tool of de collectie van sites, maar dit is slechts een klein deel van het volledige World Wide Web. De search-tool geeft dan hypertext links terug met URLs naar andere pagina's. Als je op deze links klikt dan bekom je documenten, foto's, geluid,... van individuele servers verspreid over de wereld.

Het is dus niet mogelijk om op het gehele web te zoeken ook al zijn er search tools die dit beweren.

Er bestaan verschillende categorieën van search tools:

1. Zoekmachines (& Meta-zoekmachines)

Enkele karakteristieken:

- Zoeken aan de hand van sleutelwoorden (die voorkomen in de pagina)
- Je krijgt de resultaten niet terug in de vorm van categorieën waarin je kan browsen.
- Bereik: van kleine zoekmachines tot zoekmachines die tot 90% van het indexeerbare web beslaan
- Resultaat krijg je terug in de vorm van URL's.
- Meta-zoekmachines zoeken vlug en oppervlakkig tegelijkertijd op verschillende individuele zoekmachines. Ze geven per zoekmachines ongeveer 10% van de resultaten terug.

Voorbeelden:

- Zoekmachines:
 - www.mamma.com
 - www.ixquick.com

- Meta-zoekmachines:
 - www.google.com
 - www.ask-jeeves.com
 - www.altavista.com

2. Onderwerp directories (= Subject Directories)

Enkele karakteristieken:

- Sites geselecteerd door mensen
- Meestal zorgvuldig geëvalueerd en up-to date gehouden, maar niet altijd (vooral als het over zeer algemene en grote subject directories gaat)
- Sleutelwoorden waarop je zoekt moeten niet zo specifiek zijn als bij zoekmachines omdat de woorden niet moeten overeenkomen met de woorden in het document
- Meestal geordend in hiërarchische categorieën
- Soms krijg je nog bijkomende informatie over de pagina
- Je kan browsen in de directories gebruik makend van algemene termen

Enkele voorbeelden:

- <http://www.yahoo.com/>
- <http://www.about.com/>

3. Gespecialiseerde databanken ("The Invisible Web")

The Invisible Web:

The visible web zijn de resultaatpagina's die je verkrijgt bij algemene zoekmachines en bij bijna alle subject directories. Het onzichtbare web is wat je niet als resultaat kan terug krijgen als resultaat:

- Bevraagbare databanken (waarin je gegevens kan opzoeken).
- Sommige types van documenten: niet alle documenten worden in de databank van zoekmachines opgeslagen, de databanken zijn al enorm groot en duur.

Enkele karakteristieken:

- Het Web geeft toegang om te zoeken in databanken die opgeslagen zijn in servers
- Zeer veel verschillende onderwerpen: bijvoorbeeld een databank van alle vliegtuigrampen uit de geschiedenis,...

De zoekmachine Google



(deze informatie komt hoofdzakelijk van op de website www.google.com)

Algemeen:

Basis-zoekbewerking

Om een zoekbewerking op te geven, typ je gewoon enkele woorden en druk je op de 'Enter'-toets (of klik je op de knop Google zoeken) om een lijst met relevante resultaten te krijgen.

Google zoekt enkel naar pagina's die precies overeenstemmen met je zoektermen. Als je niet het gewenste resultaat krijgt, probeer je misschien best je zoekbewerking anders te formuleren.

Voorbeeld: als je zoekt op "goedkope vliegtuigtickets" en "goedkope luchtvaarttickets", krijg je verschillende resultaten.

Automatische "AND" zoekbewerkingen

Google voegt automatisch "AND" toe tussen de woorden die je ingeeft, zodat enkel die pagina's worden gevonden waarin *al* je zoektermen voorkomen. Om een zoekbewerking verder te beperken, kan je gewoon meer termen opnemen. Google geeft ook de voorkeur aan pagina's waarin verwante zoektermen dicht bij elkaar staan.

Voorbeeld: Als je een vakantie in Griekenland plant, typ je gewoon

Vakantie Griekenland

Google zoeken

in het zoekveld en klik je op Google zoeken. Als resultaat krijg je alle pagina's waarin beide woorden voorkomen.

Beperken van zoekresultaten.

Soms gaat een zoekbewerking naar het juiste gebied, maar zijn er te veel resultaten. Om het aantal zoekresultaten te beperken, kan je een nieuwe zoekbewerking opgeven die enkel "zoekt" binnen de URL's die de eerste zoekbewerking heeft opgeleverd. Dit wordt "een zoekbewerking verfijnen" of "zoeken binnen de huidige zoekresultaten" genoemd.

Google maakt dit proces eenvoudig. Google levert namelijk enkel webpagina's op die alle woorden in je zoekbewerking omvatten. Je moet dus enkel meer zoekwoorden toe voegen aan de termen die u reeds heeft ingegeven. Deze nieuwe zoekbewerking levert dan een selectie op uit de pagina's van je oorspronkelijke "te ruime" zoekbewerking.

Voorbeeld: Als je bijvoorbeeld naar Athene zelf wil gaan tijdens je vakantie.

Vakantie Griekenland Athene

Zoekresultaten verfijnen

Sleutelwoorden: vorm en gebruik van bepaalde tekens

Hoofdletters

De zoekbewerkingen van Google houden geen rekening met hoofd- of kleine letters. Alle letters, ongeacht hoe je ze typt, worden beschouwd als kleine letters.

Voorbeeld: als je zoekt op "napoleon", "Napoleon" en "nApOIEoN", krijgt je dezelfde resultaten.

OR en AND

Google voert automatisch een "AND" zoekbewerking uit voor alle zoektermen. *Google ondersteunt de logische operator "OR" niet. Sommige zoekrobots leveren pagina's op die slechts enkele van de ingegeven zoektermen bevatten, en soms krijgen ze zelfs de voorkeur op pagina's die alle termen bevatten. Google levert enkel pagina's op die alle termen bevatten.*

Omdat Google geen "OR" zoekbewerkingen ondersteunt, kan je niet opgeven dat Google pagina's moet aanvaarden met woord A of woord B. Je kan de zoekbewerking twee keer opgeven, één keer met woord A en één keer met woord B als je de eerste keer het gewenste resultaat niet hebt gevonden.

Opmerking: je moet altijd opletten of de informatie die op een bepaalde site staat nog up-to-date is. Blijkbaar ondersteunt Google nu wel al de OR-operator. Het is dus mogelijk om te zoeken naar pagina's met woord A of woord B.

Gebruik van stamvormen en jokers

Een joker is een speciaal teken dat in de plaats staat van 1 of meerdere willekeurige tekens.

Voorbeeld: als je Griek*land als zoekcriterium ingeeft, zal je resultaten terugkrijgen met het woord Griek en het woord land maar niet Griekenland. Met een stamvorm wordt bedoeld dat het woord dat je ingeeft een deel van een ander woord kan zijn.

Voorbeeld: als je winter als zoekcriterium ingeeft dan krijg je bijvoorbeeld geen resultaten terug met enkel het woord wintersport.

Zoekbewerkingen met uitdrukkingen

Met Google kan je uitdrukkingen zoeken door aanhalingstekens te gebruiken. Woorden tussen dubbele aanhalingstekens ("zoals hier") verschijnen samen in alle gevonden documenten (tenzij het stopwoorden zijn, speciale vaak gebruikte woorden die een "+" teken nodig hebben -- zie hierna). Zoeken op uitdrukkingen met behulp van aanhalingstekens is interessant als je bekende uitspraken of specifieke namen zoekt.

Bepaalde karakters worden beschouwd als verbindingsteken voor uitdrukkingen. Google herkent koppelttekens, schuine strepen, punten,

gelijkheidstekens en apostrofs als verbindingstekens voor uitdrukkingen. Dergelijke verbindingstekens hebben dezelfde functie als aanhalingstekens; zo wordt up-to-date beschouwd als een uitdrukking, hoewel de drie woorden niet tussen aanhalingstekens staan.

Stopwoorden

Google negeert vaak gebruikte woorden en karakters, die we stopwoorden noemen. Google negeert automatisch termen zoals "*http*" en "*com*", alsook bepaalde losstaande cijfers en letters, omdat die termen zelden kunnen helpen om een zoekbewerking te verfijnen en ze het zoekwerk sterk kunnen vertragen.

Gebruik het teken "+" om stopwoorden in uw zoekbewerking op te nemen. Vergeet niet een spatie te plaatsen voor het "+" teken. U kunt ook het teken "+" opnemen als u naar uitdrukkingen zoekt.

Als u zoekt naar Star Wars, Episode I:

Episode +I	Google zoeken
------------	---------------

Woorden uitsluiten uit uw zoekbewerking

Soms is het handig om bepaalde woorden of uitdrukkingen uit te sluiten uit een zoekbewerking; je wenst alle relevante resultaten, behalve resultaten met een bepaald woord of een bepaalde uitdrukking. Google ondersteunt deze "negatieve" functionaliteit met het min-teken ("-"). Gebruik het min-teken om een term uit je zoekbewerking uit te sluiten. Vergeet niet een spatie te plaatsen voor het min-teken, zodat Google alle pagina's met dat woord kan negeren.

Voorbeeld: Als je wil zoeken naar president, maar Bush is niet je grootste vriend dan geef je het volgende in:

president -Bush	Google zoeken
-----------------	---------------

Speciale zoekfuncties

Als bepaalde woorden worden gevolgd door een dubbele punt, hebben ze een speciale betekenis voor Google.

link:

De vraag *link: <url>* toont alle terugkoppelingen voor een bepaalde URL -- met andere woorden: de pagina's die naar die URL verwijzen.

Voorbeeld:

link:www.google.com	Google Zoeken
---------------------	---------------

Dit toont alle pagina's die verwijzen naar de home page van Google. Je kan een zoekbewerking met link: niet combineren met een gewone zoekbewerking met sleutelwoorden.

cache:

Dit doe hetzelfde als [in cache](#).

cache:www.google.com	Google Zoeken
----------------------	---------------

related:

Dit doet hetzelfde als [GoogleScout](#).

related:www.google.com	Google Zoeken
------------------------	---------------

info:

Toont de info die Google heeft over een bepaalde pagina.

info:www.google.com	Google Zoeken
---------------------	---------------

spell:

Als je het volgende intikt dan zal Google de spelling controleren en eventueel enkele alternatieven voorstellen.

spell:corect my speeling	Google Zoeken
--------------------------	---------------

stocks: (niet belangrijk)

site:

Google beperkt de resultaten tot deze in het opgegeven domein.

help site:com	Google Zoeken
---------------	---------------

Hier wordt als resultaat alle pagina's binnen het com domein over help.

allintitle:

Alleen de resultaten met alle sleutelwoorden in de titel worden weergegeven.

allintitle: google search	Google Zoeken
---------------------------	---------------

intitle:

Het woord dat na intitle: voorkomt staat in de titel. Er kunnen nog andere sleutelwoorden voorkomen, maar deze moeten niet noodzakelijk in de titel voorkomen.

allinurl:

Het resultaat zijn alle URL's die alle sleutelwoorden bevatten.

allinurl: google search	Google Zoeken
-------------------------	---------------

inurl:

Het woord juist na inurl: moet in de URL voorkomen. Er kunnen nog andere sleutelwoorden voorkomen, maar deze moeten niet noodzakelijk in de URL voorkomen.

Zoekresultaten interpreteren:



A. Zoekveld

Om een zoekbewerking op te geven, typ je gewoon enkele sleutelwoorden en klik je op de knop Google zoeken, om een lijst met relevante resultaten te krijgen. De sleutelwoorden die je ingeeft zullen allemaal voorkomen in de resultaten. Als je wil dat bepaalde woorden samen voorkomen dan kan je deze tussen aanhalingstekens plaatsen.

B. Knop Google zoeken

Klik op deze knop om een nieuwe zoekbewerking in te dienen. U kunt ook een zoekbewerking indienen door op de 'Enter'-toets te drukken.

C. Ik ga d'r voor – Ik grijp mijn kans

De knop "Ik grijp mijn kans™" brengt je rechtstreeks naar de eerste webpagina die Google voor de zoekbewerking heeft gevonden. Je ziet de andere zoekresultaten niet.

Voorbeeld: Als je de home page van het weekblad Humo wenst te vinden, typ je gewoon het volgende

in het zoekveld en klik je op de knop "Ik ga d'r voor". Google brengt je rechtstreeks naar "www.humo.be", de officiële home page van Humo.

D. Resultatenmenu

Je kan het aantal resultaten per pagina instellen.

Dit vind je bij Voorkeuren, waar je nog andere instellingen kan wijzigen.

E. Paginatitel

De eerste regel van het resultaat is de titel van de gevonden webpagina. Soms staat er geen titel maar wel een URL. Dit betekent dat we de gevonden webpagina nog niet geïndexeerd hebben en dus ook de titel nog niet kennen. Het kan ook betekenen dat de auteur geen titel aan zijn pagina heeft gegeven. We weten echter wel dat het een goede treffer betreft, want andere webpagina's -- die we wel geïndexeerd hebben -- beschikken over koppelingen naar deze pagina. Als de tekst bij deze koppelingen beantwoordt aan uw zoekbewerking, kunnen we de pagina als zoekresultaat tonen hoewel hij nog niet geïndexeerd is.

F. Google en RealNames

Het RN (RealNames) merkteken dat soms verschijnt in de pagina met Google zoekresultaten, wijst op een nieuw partnership om de Google-zoekfunctie te verbeteren en de integriteit van een Google-zoekbewerking in stand te houden.

Internet-sleutelwoorden zijn merk-, product-, dienst- of bedrijfsnamen die specifiek naar de website van de eigenaar verwijzen. Voorbeeld: Jeep Grand Cherokee^{RN} brengt u naar de Grand Cherokee pagina in de officiële site van Jeep.

Als het Internet-sleutelwoord overeenkomt met het eerste resultaat van Google, verschijnen het Internet-sleutelwoord en RN op het einde van de titel. Als het Internet-sleutelwoord verschillend is van het eerste resultaat van Google, verschijnt het boven de Google-resultaten.

G. Tekst onder de titel

Deze tekst geeft een samenvatting van de gevonden resultaatpagina. De tekst is niet noodzakelijk de eerste tekst op de pagina, maar een fragment van de pagina met één of meer door u opgegeven zoektermen in vet lettertype.

H. Statistiekregel

Deze regel geeft het aantal resultaten van uw zoekbewerking en de tijd die nodig was om uw zoekbewerking uit te voeren.

I. Ingesprongen resultaat

Als Google verschillende resultaten van dezelfde website vindt, wordt het meest relevante resultaat bovenaan geplaatst en worden de andere relevante pagina's van dezelfde site eronder geplaatst, maar

ingesprongen.

J. URL van resultaat

Dit is het webadres van het gevonden resultaat.

K. Gecachte koppeling

Achter de URL staat een gecachte koppeling. Als u op deze koppeling klikt, ziet u de inhoud van de webpagina zoals hij het laatst werd geïndexeerd. Voor websites met snel wisselende inhoud, zoals nieuwssites, kan de *actuele* inhoud van een webpagina niet goed overeenstemmen met uw zoekbewerking. De *gecachte* inhoud is de inhoud die Google gebruikt om te bepalen of deze pagina goed overeenstemt met uw zoekbewerking. De gecachte koppeling ontbreekt voor sites die niet geïndexeerd zijn en voor sites wiens eigenaars ons hebben gevraagd de gecachte inhoud te verwijderen.

Als u op de gecachte koppeling klikt, ziet u bovenaan de resulterende pagina HTTP-kopinformatie. De informatie kan op zichzelf al interessant zijn -- de datums kunnen aangeven hoe oud de gecachte inhoud is -- maar, wat belangrijker is, hierdoor ziet u dat het om een gecachte versie van de pagina gaat en niet om de pagina zelf.

Als de koppeling op de site u om een of andere reden niet verbindt met de huidige pagina, kunt u de gecachte versie opvragen en eventueel daar de gewenste informatie vinden.

L. Omvang

Het volgende getal op de regel is de omvang van het tekstgedeelte van de webpagina. Deze informatie wordt niet opgegeven voor sites die nog niet geïndexeerd zijn.

M. GoogleScout – Gelijkwaardige pagina's

Als u voor een bepaald resultaat op de koppeling GoogleScout klikt, doorzoekt Google automatisch het web naar pagina's die verwant zijn met dit resultaat. Normaal vindt GoogleScout een tiental pagina's van hoge kwaliteit. Bij het zoeken naar verwante pagina's probeert GoogleScout pagina's te vinden met dezelfde graad van algemeenheid. Voorbeeld: Als de startpagina de home page van een universiteit is, levert GoogleScout de home pages van andere universiteiten op. Maar als de startpagina de computerafdeling van de universiteit is, vindt GoogleScout verwante computerafdelingen, niet verwante universiteiten. GoogleScout kan op verschillende manieren worden gebruikt. Als u de inhoud van een bepaalde site interessant vindt, maar nog meer informatie wenst, kan GoogleScout sites met gelijkaardige inhoud zoeken waarmee u misschien niet vertrouwd bent. Als u productinformatie zoekt, vindt GoogleScout informatie van concurrenten, zodat u een gefundeerde keuze kunt maken. Als u belangstelt in een bepaald onderzoeksdomein, kan GoogleScout u helpen om zeer snel een groot aantal resources te

zoeken, zonder u af te vragen welke sleutelwoorden voor die sites effectief werden gebruikt.

GoogleScout vindt verwante pagina's voor miljoenen webpagina's. Hoe meer een pagina gespecialiseerd is, hoe kleiner het aantal resultaten dat GoogleScout voor u kan vinden. Voorbeeld: GoogleScout kan misschien geen verwante pagina's voor uw persoonlijke home page vinden als er onvoldoende informatie is om andere pagina's betrouwbaar met uw pagina te verbinden. Als bedrijven verschillende URL's voor hun pagina's gebruiken (zoals bedrijf.com en www.bedrijf.com), beschikt GoogleScout soms over heel weinig informatie voor één URL, maar over veel informatie voor de andere URL. Algemeen werkt GoogleScout goed met de meeste webpagina's.

Verschillende mogelijkheden van Google

Zoeken naar documenten

Zoeken naar images

...

Geavanceerd zoeken

Bijna alle mogelijkheden die je met geavanceerd zoeken kan, kan je ook via de speciale zoekfuncties.

Voorkeuren

Hier kan je je persoonlijke instellingen ingeven zoals taal, aantal resultaten per pagina,...

Er wordt dan een cookie naar je computer gestuurd met je voorkeuren (tenzij je in je browser hebt ingesteld dat je geen cookies wil). De volgende keer dat je naar die pagina gaat, zijn je persoonlijke voorkeuren nog ingesteld. Dit komt omdat de webserver (van google) de cookie opvraagt. Je browser zal dit dan opzoeken en de informatie naar de webserver sturen. Zo blijven je persoonlijke instellingen bewaard.