



Momenteel is er een overvloed aan zoekinstrumenten om het Web te doorzoeken. Aarzelt u niet om een zelfde zoekopdracht te laten uitvoeren door verschillende zoekinstrumenten en de relevantie van de resultaten te vergelijken! Samen met een correcte formulering van de zoekopdracht is dat nog altijd de sleutel van het succes.

r.lassaux@iec-iab.be

“ EVEN GEDULD... BEZIG MET ZOEKEN ”

THE DEEP BLUE ...

Iedereen weet dat het Web bestaat uit miljoenen en miljoenen inlichtingen en diensten, verdeeld over nagenoeg evenveel sites. Minder duidelijk is hoe groot het Web precies wel is: eind 2000 bestaat deze uitzonderlijke informatiebron uit niet minder dan twee miljard webpagina's, waarvan de helft overigens niet toegankelijk zou zijn. Als u dan nog weet dat er elke seconde 80 pagina's bijkomen (elke dag 7 miljoen), zal het u niet verwonderen dat geen enkel zoekinstrument op dit moment in staat is alle woorden en alle uitdrukkingen op alle pagina's te inventariseren. Sommige instrumenten pretenderen openlijk vanaf 2001 heel het Web te beslaan. Wij vragen ons af of dat eigenlijk wel nodig is. Immers, om snel die witte raaf, de site of pagina die perfect beantwoordt aan uw verwachtingen, te vinden, moet u in de eerste plaats weten waar u moet zoeken. En dat is precies de bestaansreden van de zoekinstrumenten waarover dit artikel gaat. We kennen ze onder verschillende namen: gidsen, zoekmachines, discussiegroepen en gespecialiseerde sites. Ze hebben allemaal hun sterke en hun zwakke punten en de eerste taak van alle nieuwe surfers bestaat er uiteraard in uit te vinden welk(e) zoekinstrument(en) geschikt zijn onder welke omstandigheden.

Om een keuze te kunnen maken, moet u weten hoe de verschillende zoekinstrumenten functioneren, welke informatie ze kunnen zoeken en welke instrumenten het

krachtigst zijn. In dit artikel willen wij een antwoord geven op al deze vragen voor elk zoekinstrument. Eerst geven we een overzicht van de beschikbare instrumenten en vervolgens vullen we onze conclusie aan met enkele praktische tips die u in uw achterhoofd moet houden bij uw omzwervingen op het Web.



een zoekopdracht invoert, gaat de zoekmachine in haar databank integraal op zoek naar de webpagina's waarvan de tekstuele inhoud overeenstemt met de trefwoorden die u hebt ingevoerd. Zodra ze bepaald heeft welke pagina's de zoekterm bevatten, klasseert de zoekmachine de pagina's volgens hun relevantie, afhankelijk van een volgorde en een algoritme (op basis van bepaalde sorteercriteria) die voor elke zoekmachine anders is. Later overloopt de zoekmachine de geïndexeerde pagina's regelmatig opnieuw - min of meer frequent - om er een recentere versie van op te slaan. Dat noemt men de "refreshing"-fase.

Er zijn honderden zoekmachines en dit werkingsprincipe geldt voor allemaal, maar dat is nagenoeg het enige wat ze gemeenschappelijk hebben. Om de relevantie te waarborgen passen de zoekmachines steeds meer middeljes toe om het Web te klasseren, te ordenen en te indexeren. Hieronder vindt u een kleine selectie, met argumenten ter ondersteuning.

Zoals te verwachten zijn de belangrijkste zoekmachines Amerikaans. De onbetwistbare ster in de V.S., **Google** (een databank van 560 miljoen pagina's met als doelstelling meer dan een miljard pagina's vanaf 2001), dankt haar succes

aan een vernieuwend zoekstelsel: de populariteitsindex¹ van de pagina's. Een site wordt niet alleen geselecteerd op basis van de informatie die hij bevat, maar ook op basis van het aantal linken naar de desbetreffende site op het Internet. Hoe meer er naar een pagina verwezen wordt op websites, hoe 'populairder' de pagina is, hoe interessanter ze gevonden wordt en hoe hoger de plaats die ze krijgt bij de zoekresultaten. Verder noemen we nog **Altavista** (300 miljoen geïndexeerde pagina's)

dat een bijzonder snelle nieuwe zoekmachine **Raging** heeft. Uiteraard mogen we ook **Hotbot** niet vergeten. Deze zoekmachine, die door velen beschouwd wordt als een van de beste algemene zoekmachines van het moment op het Internet, biedt u de mogelijkheid alleen te zoeken naar pagina's die niet ouder zijn dan een bepaalde periode: een of twee weken, een maand enz. Hoewel slechts 10 % van de webpagina's in het Frans zijn opgesteld, doet het ons plezier vast te stellen dat de Franstalige zoekmachines momenteel niet veel meer

moeten onderdoen voor hun Engelstalige tegenhangers. De onbetwistbare leiders in Frankrijk zijn **Voila**, (150 miljoen geïndexeerde pagina's) en **Altavista France**. Maar andere, zoals **Ecila**, **Excite**, **InfoSeek**, **Lokace**, **OpenText**, **SuperSearch**, **Caloweb** en uiteraard **Lycos** behalen ook zeer goede resultaten. Dat geldt ook voor de beroemde Belgische zoekmachines **Advalvas** en **Webbel**.

Algemene zoekmachines zijn nauwkeuriger dan gidsen omdat ze de inhoud van de webpagina's onder de loep nemen, maar u mag zich niet beperken tot één zoekinstrument. Als uw zoekopdracht niet correct geformuleerd is of te algemeen, zult u bovendien bedolven worden onder de zoekresultaten. Als u gerichte informatie zoekt op het Internet, kunt u beter gebruik maken van 'metazoeckmachines' of gespecialiseerde zoekmachines.

METAZOEKMACHINES

Met dit vreemd woord, dat niet zou misstaan in de betere fantastische film, benoemt men een software die in staat is meerdere zoekmachines tegelijkertijd naar een term te laten zoeken en de resultaten gecombineerd weer te geven.

De bekendste metazoeckmachine is **Copernic**² 2000. De basisversie is freeware die gratis gedownload kan worden (2,3 Mb) op de site van Copernic Technologies. Voor de Engelstalige sites zult u op **all4one**, **mamma**, **infind** en **dogpile** de resultaten vinden die u gehoopt had. Ten slotte verdient ook **SoftCrawler** onze aandacht. Dit waardevol hulpprogramma kan de catalogi doorzoeken van tien belangrijke downloadsites, waaronder **Shareware**, **Download**, **SoftSeek**, **FileMine**, **Zdnet**, **Hotfiles**, **Tucows**, **Alberts**, **LinuxFiles**, **32bit** en **Lycos**.

Op dit moment hebben de metazoeckmachines nog maar twee zwakke punten: ze zijn relatief traag en ze laten geen

gerichte zoekopdrachten toe doordat de syntaxis van de verschillende zoekmachines niet compatibel is.

GESPECIALISEERDE TOOLS

Doordat ze informatie over een specifiek vakdomein (e-commerce, geneeskunde, toerisme, wetgeving enz.) concentreren, zijn deze sites de ontmoetingsplaats voor liefhebbers van gespecialiseerde informatie.



De gespecialiseerde sites die wij genoemd hebben in onze vorige rubrieken zijn hier al een duidelijk voorbeeld van: euro, fiscaliteit, vennootschapsrecht. Daarbij komt nog de zeer recente portaalsite voor het Belgische recht (www.droitbelge.net). Tot deze categorie behoren ook de sites die op basis van een naam e-mailadressen inventariseren (**Four 11** of **WhoWhere**) of telefoonnummers, zoals **Infobel**. Er zijn er nog zoveel andere van allerlei aard³, maar degene die zeker een vermelding verdienen zijn: **Scour**, een zeer degelijke gespecialiseerde zoekmachine voor multimedia-informatie, **Ftpsearch**, de huidige standaard voor het vinden van het stuurprogramma voor randapparatuur, het bestand, de freeware of shareware die u nodig hebt, en **Babylon**, een uitstekend woordenboek Frans-Engels waarvan de versie 3.0 gratis te downloaden is op de site van de uitgeverij

Het voordeel van deze formule is duidelijk. Behalve een snelle toegang tot de informatie, sluit de nauwkeurige afbakening van het zoekdomein ook verwarring met homoniemen uit. Maar de medaille heeft een keerzijde. Als men het zoekgebied te veel beperkt, is de kans reëel dat informatie die perfect in overeenstemming is met wat u zoekt, buiten beschouwing gelaten wordt.

NIEUWSGROEPEN, DISCUSSIEGROEPEN

U kunt uw vraag ook meteen stellen aan andere surfers die uw interesses delen.

Dergelijke groepen zijn toegankelijk via de meeste algemene (**Yahoo**, **Lycos**...) of gespecialiseerde gidsen (**Forumone**, **Forum.multimedia**, **Oragora**), of via zoekmachines als **InfoScan** en vooral **Al-News**. Er is dus niets eenvoudiger dan bij de discussiegroepen te gaan zoeken in de databank van miljoenen ingezonden berichten.

Dit mag dan heel aantrekkelijk klinken, maar het systeem zelf houdt ook beperkingen in waarvan de gebruikers zich bewust moeten zijn. Hoewel u er vaak de beste informatie vindt - vaak zullen specialisten uit de sector reageren - kunt u ook slecht geïnformeerd worden - het staat immers iedereen vrij te publiceren op het Internet wat hij wil. Het meedelen van uw e-mailadres houdt ook risico's in. Zo zet u de deur open voor *spam* (reclameboodschappen), *mail bombing* (massale verzending van berichten), en niet te vergeten virussen. Om zo discreet mogelijk te blijven, kunnen wij u niet genoeg op het hart drukken een tweede

e-mailadres te nemen dat u alleen daarvoor gebruikt, of zelfs om uw vraag te versturen via sites die in principe uw anonimiteit garanderen, zoals (**anonymizer**).

STAND VAN ZAKEN...

Het Web is exponentieel gegroeid en omvat vandaag meer informatie dan ooit voorheen. De steeds gebruiksvriendelijkere zoekinstrumenten reiken altijd maar nieuwe snuffjes aan om het internet te klasseren, te ordenen en te indexereren, maar paradoxaal genoeg is het nooit moeilijker geweest om er relevante informatie in terug te vinden. Zomaar willekeurig zoeken op het Internet is als zoeken naar een naald in een hooiberg. Maar vaak zorgt een klein beetje knowhow voor mirakels. Aansluitend bij wat op de voorgaande pagina's werd uiteengezet, geven wij u hieronder tien tips die u op weg moeten helpen voor een efficiënte opzoeking.



① Alle zoekinstrumenten leveren resultaten op die vaak betrouwbaar zijn, maar men moet wel de spelregels kennen, met andere woorden hun potentieel en vooral hun grenzen. Ze hebben allemaal hun sterke en hun zwakke punten. Gebruikt u gerust het ene instrument na het andere voor eenzelfde opzoeking en vergelijk hun resultaten. Een gecombineerd gebruik brengt u vaak waar u wezen moet ...

② De zoekmachines kunnen tot de doeltreffendste instrumenten worden gerekend. Net als de beste wijnen

worden ze nog beter met de jaren, en de nieuwkomers lopen over van de originele verbeteringen⁴ om toch maar een plaatsje onder de zon te veroveren.

De ideale zoekmachine beschikt over de volgende kenmerken: ze is zeer snel, in staat om alles te vinden, toont de resultaten op een geschikte manier, heeft een duidelijke interface die is uitgerust met originele en praktische functionaliteiten (geavanceerde opzoeking, semantische analyse, ...), verzendt zoekrapporten per elektronische post, is voorzien van een waarschuwingsfunctie die veranderingen of nieuwe documenten over een bepaald thema detecteert, is beveiligd, soepel, evolutief en, - kers op de taart, - in staat blanco links te herkennen, zodat u het raadselachtige maar ergerlijke bericht "error 404- page not found " kan vermijden...

Maar laat ons deze droom (die niet zo ver meer verwijderd is) terzijde schuiven en de realiteit van elke dag bekijken. Dit ideale instrument, dat het beste zal nemen van alle technieken om zo doeltreffend mogelijk aan de slag te gaan, leeft momenteel alleen in de rijke verbeelding van de ontwerpers; toch kunnen we een aantal kenmerken ervan al aantreffen in de huidige machines. Als u tot het zover is toch zeker wil dat u het beste haalt uit het Web in één enkele beweging, dan zijn de metamachines voor u de ideale gezelschap.

③ Naast deze grote algemene machines voeren gespecialiseerde machines u mee naar de downloadsites, de discussieforums, de elektronische adressen, zelfs de online gidsen. Een bezoekje eraan is gewoonlijk de omweg waard.

④ Hoewel de formule nog weinig wordt gebruikt voor het opzoeken van informatie, kunnen de forums en de discussiegroepen niet systematisch worden opzijgeschoven. Ze leveren vaak de slechtste maar anderzijds ook de beste resultaten op.



⑤ Tenzij u het Web constant zit uit te vlooien, is het actualiseren van uw kennis geen sinecure. Voortaan hoeft dit echter geen probleem meer te zijn, aangezien verschillende sites zich erop hebben toegelegd de actualiteit en de evolutie van de zoekmotoren op de voet te volgen. Het nirwana ter zake is de Franse site Abondance (www.abondance.fr), die zeer regelmatig wordt geüpdatet en die, naast gedetailleerde informatie over de zoekmachines en hun werking, vaak ook vergelijkingen van de verschillende tools voorstelt. Heel wat gedetailleerde informatie over de prestaties en kostbaar advies over de beste strategieën vindt u ook, met één enkele muisklik, op searchenginesshowdown.com, in het Engels.

⑥ Als u tijd wil winnen, dan is het een goed idee uw navigator te configureren. Het is heel eenvoudig: u kiest gewoon de tool die het best geschikt is als startpagina bij het opstarten van uw navigator. In Microsoft Explorer (versie 5) klikt u op Extra/Internet-opties en vervolgens kiest u het tabblad Algemeen. Typ vervolgens het adres van het portal in het vakje Startpagina. Met Netscape Navigator (4. 7) klikt u op Bewerken/Voorkeuren, vervolgens kiest u Navigator. Typ het adres van uw portal in het tekstvak Startpagina. Gebeurd.

⑦ Indien u vaak dezelfde site bezoekt, dan kan u zeker uw voordeel doen met de applicaties om hele sites te downloaden, zoals [Memo web 3 flash](#) of [webWhacker](#)... Met deze gespecialiseerde software brengt u in een handomdraai een site volledig of gedeeltelijk onder op uw harde schijf. U hoeft dan geen communicatiekosten meer te betalen, maar uiteraard is deze functie niet geschikt voor sites die regelmatig worden bijgewerkt.

⑧ Als u na een paar minuten zoeken op het web oevervallen wordt door moedeloosheid, dan is dit heel begrijpelijk. Laat de moede echter niet helemaal zakken. Speel uw vraag door aan een expert. Het online raadplegen van een site-expert is momenteel erg in zwang: deze zoekmethode houdt in dat u een exact antwoord te krijgt op een exacte vraag, en bovendien binnen een welbepaalde termijn. Mits betaling geven sites als www.question.fr, www.paypersearch.com, www.woonoz.com of, internationaal, www.exp.com, www.askme.com en www.veille.com afdoende antwoorden op uw vragen in gevarieerde domeinen als gezondheid, informatica of financiën. Bovendien is het mogelijk automatisch op de hoogte te worden gehouden van de evolutie van bepaalde informatie, zonder dat u zelf op zoek hoeft te gaan. Dit kan op twee manieren. De eerste heet Spyonit (www.spyonit.com) en is een zoekmachine die het web afspeurt en u (via e-mail, gsm, instant messaging, ...) op de hoogte brengt van nieuwtjes over het onderwerp van uw opzoeking. U kan ook een abonnement nemen op een distributielijst. Er bestaan tal van categorieën in het Frans (www.franco-phones.listes.com) of in het Engels (www.liszt.com). U ontvangt dan alle informatie via uw mailbox.

⑨ Houd er te allen tijde rekening mee dat de zoektools geen garanties bieden voor de correctheid van de gevonden informatie. Het is dus nog altijd beter grondig te controleren.

⑩ De sleutel van het succes ligt echter evenzeer, of vooral, in een optimale formulering van de vraag. Hiervoor kun-

nen wij u geen mirakels aan de hand doen, maar wel een aantal praktische trucs en kneepjes om het zoekveld te beperken, zoals het gebruik van de geavanceerde zoekfuncties die alle grote machines aanbieden of het gebruik van de Booleaanse operatoren. Maar daarover leest u meer in ons volgende artikel.

BINNENSTEBUITEN GEKEERD VAN HET WEB ⁵

GIDSEN	Algemeen	Thematisch
	Yahoo! International www.yahoo.com Nomade www.nomade.fr Lycos www.lycos.fr Msn www.msn.fr Wanadoo www.wanadoo.fr Voila www.voila.fr	7Alpha7 www.7alpha.com/ Enfin www.enfin.com/ Search Engine Guide www.searchengineguide.com/
MACHINES	Algemeen	Metamachines
	Altavista www.altavista.fr ou www.altavista.com Lycos www.lycos.fr ou www.lycos.com Caloweb www.caloweb.com Webcheck www.webcheck.com Lokace www.lokace.fr Hotbot www.hotbot.com Google www.google.com Go www.go.com Webcrawler www.webcrawler.com Infoseek www.infoseek.com Alltheweb, allthe time http://www.alltheweb.com	Copernic 2000 www.copernic.com.fr SoftCrawler www.data-wizard.de/SoftCrawler ALL4one www.all4one.com Mamma www.mamma.com Infind www.infind.com Dogpile www.dogpile.com1
GESPECIALISEERDE TOOLS	Question www.question.fr , Paypersearch www.paypersearch.com Expert www.exp.com , Veille www.veille.com	Belgisch recht www.droitbelge.net Scour www.scour.com Ftpsearch www.ftpsearch.com Searchability www.searchability.com Babylon www.babylon.com

- Deze indexering van het Web door het aantal en het belang van de linken te analyseren, heeft echter het nadeel dat recente en/of zeer gespecialiseerde pagina's uit de boot vallen omdat er maar weinig sites daarnaar verwijzen.
- De nieuwe interface omvat 55 zoekmachines (tegenover 40 vroeger), verdeeld over zes domeinen: heel het Web, discussiegroepen, elektronische adresboeken, verkoop van boeken, verkoop van software en verkoop van hardware. Zodra de zoekopdracht uitgevoerd is, zal de machine de documenten sorteren, de dubbels verwijderen, controleren of de pagina's nog bestaan en de zoektermen onderstrepen. De machine bewaart een historiek van de zoekopdrachten en kan deze enige tijd later verfijnen of actualiseren. Bovendien voegt het een opdracht toe aan het snelmenu van de rechtermuisknop, zodat u een zoekopdracht kunt starten door met de muis op een woord te klikken. Voor iedere zoekopdracht downloadt de software automatisch de laatste nieuwe zoekmodules om zichzelf te actualiseren.
- Veel zoekmachines geven een overzicht van de gespecialiseerde sites met een korte beschrijving van enkele regels van elke site: *Beaucoup, Internet Sleuth, 7 Alpha IDF, Search Power, The Ultimate Search Page, Special search engines, FinderSeeker, Internet Search Engines, BIG Search Engine Index.*
- Sommige baanbrekende zoekmachines zoals *Excalibur* aanvaarden voortaan ook onnauwkeurige verzoeken: ze zoeken niet alleen een antwoord voor de exacte ingevoerde tekenreeksen, maar voor alle tekenreeksen die in de buurt komen. Om de pagina's te vinden die zo goed mogelijk

- beantwoorden aan de verzoeken, gebruikt Caloweb eveneens geavanceerde technieken van tekstovereenkomst. Zonder af te wijken van de regel brengt de zoekmachine Webcheck een interessant idee binnen ieders bereik: het zoeken volgens concept. In de plaats van te vertrouwen op het aantal herhalingen van een of meer sleutelwoorden, doorzoekt WebCheck zo'n 500 miljoen geïndexeerde pagina's om degene eruit te halen die verwijzen naar het thema dat rechtstreeks verband houdt met de vraag. We vermelden ook Cobrain, maar vooral Oingo dat, in de plaats van te zoeken waar een woord voorkomt op een pagina, werkt met combinaties van ideeën, dankzij een woordenboek waarop de machine zich beroept. Wanneer een term wordt gevraagd, antwoordt ze met een lijst van betekenissen waaruit de gebruiker dan degene kan kiezen die het best past. Ook Copernic Summarizer kan bogen op nieuwe functies: de software vat webpagina's samen en laat de teksten achter zich. De software is heel gemakkelijk te gebruiken: u geeft hem gewoon het adres van een webpagina en hij vat ze samen, waarbij hij alleen de concepten onthoudt. De software is ingebouwd in Internet Explorer en Word. Tot slot zijn er nog tal van laboratoria die werken aan zoekmachines die foto's kunnen terugvinden zonder dat hiervoor één enkel woord moet worden gebruikt, op basis van een schets of aanwijzingen over de kleur of de textuur. De toekomst heeft dus duidelijk nog heel wat belangrijke veranderingen in petto.
- Dit artikel kan u on line raadplegen op onze Internet-site (<http://www.accountancy.be>). De dynamische links naar alle geregistreerde sites zijn geactiveerd zodat één muisklik u naar het url-adres brengt.