

# Internetgidsen en het diepe web

**Z**oekmachines zijn niet de enige manier om op internet te zoeken, gelukkig maar. Zoals in hoofdstuk 2 is besproken, doen zoekmachines geautomatiseerd hun werk. Internetgidsen daarentegen worden door mensen samengesteld. Hoewel dat soms tot bijzonder slechte resultaten leidt, heeft het ook een groot voordeel: mensen hebben inzicht in het diepe web, het deel van internet dat voor zoekmachines onvindbaar is. Kortom, dit hoofdstuk is een ode aan het meest dominante tweepotige zoogdier van de planeet Aarde.

## **U leert in dit hoofdstuk:**

*Wat het verschil is tussen goede en slechte internetgidsen*

*Enkele slechte zoekmachines van dichtbij kennen*

*Enkele goede zoekmachines van dichtbij kennen*

*Wat het diepe web is en waarom zoekmachines er geen vat op hebben*

*Hoe u het diepe web toch kunt doorzoeken*

## Goede versus slechte internetgidsen

Een kleine herhaling van hoofdstuk 2: een internetgids is een gerubriceerde indeling van door (bijna) menselijke redacteurs bezochte en geëvalueerde sites. Die moeten niet alleen beslissen welke sites ze opnemen, maar ook in welke rubriek deze vervolgens worden geplaatst. Dat geeft gelijk al aan waar de potentieel zwakke punten van een internetgids liggen: op het gebied van structuur, kwantiteit en kwaliteit.

- **Structuur** Een internetgids zonder logische rubrieksindeling is onbruikbaar. Van belang is ook hoe de redacteurs van de internetgids zich aan hun eigen structuur houden. Dat lijkt een open deur, maar veel internetgidsen vallen hier door de mand. Een heldere rubrieksindeling maken valt al niet mee, maar deze consequent handhaven, mislukt zeer regelmatig.
- **Kwantiteit** Volledig kan een internetgids nooit zijn. Dat lukt een zoekmachine niet eens. Maar een goede internetgids moet zich niet beperken tot een beperkt aantal rubrieken of de meest bekende sites voor een bepaald onderwerp. Bestaat een internetgids alleen uit sites uit de laatste categorie, dan voegt deze immers weinig toe aan wat veel researchers al weten. Het nadeel van een internetgids met weinig rubrieken is dat u een waslijst aan nuttige internetgidsen moet bijhouden om al uw onderzoeksbehoeften op te kunnen botvieren.
- **Kwaliteit** Zeker gezien de beperkingen ten aanzien van kwantiteit moeten de koppelingen die overblijven veel beter zijn dan de resultaten die een zoekmachine uitspuugt. Wie een database heeft van miljarden internetpagina's kan best wat rotte appels verdragen; wie er slechts enkele honderden of duizenden in de aanbieding heeft, moet ervoor zorgen dat zijn gehele aanbod tenminste aan enige basale kwaliteitseisen voldoet.

Uit het voorgaande blijkt al dat het maken van een goede internetgids heel veel tijd en geld kost. Een aantal van de internetgidsen dat hierna besproken wordt voldoet daaraan, andere niet.

### Startpagina – [www.pagina.nl](http://www.pagina.nl)

Een internetgids die maar aan een van de drie genoemde criteria voldoet: kwantiteit. Op Startpagina staan heel veel koppelingen, bij elkaar geharkt door vrijwilligers. Deze krijgen in ruil daarvoor de helft van de advertentie-inkomsten op hun gedeelte van de Startpagina-site. Helaas wisselt de kwaliteit van deze verschillende gedeeltes, ook wel Startpagina-dochters genoemd, onderling sterk. Sommige dochters zijn zeer commercieel, andere

bevatten daadwerkelijk goede koppelingen, wat een bezoek aan Startpagina onvoorspelbaar maakt. De rubricering is een rommeltje: meer dan eens staan sites in rubrieken waar ze niet thuishoren. Ook zijn er sites die in meerdere rubrieken prima tot hun recht zouden komen, maar slechts eenmaal worden vermeld.

### Yahoo! – [directory.yahoo.com](http://directory.yahoo.com)

Hoe je een goed concept om zeep helpt, laat Yahoo! zien. Het bedrijf dat de internetbevolking de eerste goedwerkende internetgids liet zien, heeft zijn bedenkens laten verpieteren. De gids zit vol met verouderde of niet-relevante links, de scheiding tussen commercie en redactie is mild uitgedrukt nogal vaag en wat betreft kwantiteit scoort Yahoo! ook al niet goed. Bij het schrijven van dit boek wist Yahoo! geen andere landelijke kranten te noemen dan de Telegraaf, de Volkskrant en NRC Handelsblad.



#### Gebruik de rubrieksindeling

Gebruik bij het zoeken in internetgidsen de rubrieksindeling. Redacteurs van een internetgids maken de indeling van koppelingen immers ook op basis van deze structuur. Weliswaar bevatten veel internetgidsen een zoekveld, maar deze functie werkt vaak niet optimaal.

### About.com – [www.about.com](http://www.about.com)

Zo kan het ook. De Amerikaanse site About.com wordt net als Startpagina volgeschreven door vrijwilligers die meedelen in de advertentiewinst. Het verschil is dat deze vrijwilligers aanzienlijk beter hun best doen: de About.com-vrijwilligers geven niet alleen een verzameling koppelingen naar andere sites, maar ook advies en actuele informatie. Zo had About.com na de aanslagen van 11 september 2001 binnen korte tijd uitstekende overzichten met koppelingen. Deze grondige aanpak vertaalt zich wel in een kwantitatieve beperking: Startpagina heeft meer dochters dan About.com op onderwerp gespecialiseerde internetgidsen heeft. Gebruik nooit het zoekveld op About.com, maar gebruik het rubriekssysteem: dit om terminale verwarring te voorkomen. Veel About.com-gidsen zijn overigens eenvoudig te vinden door het betreffende onderwerp voor het siteadres te zetten. De About.com-gids over Islam is bijvoorbeeld te vinden op [islam.about.com](http://islam.about.com); die over politieke humor op [politicalhumor.about.com](http://politicalhumor.about.com).

## Hoofdstuk 5 – Internetgidsen en het diepe web



**Afbeelding 5.1** Laat u niet afschrikken door het al te gelikte uiterlijk.

### DMOZ – [www.dmoz.nl](http://www.dmoz.nl)

Directory Mozilla, beter bekend als het Open Directory Project, is een goede en uitgebreide internetgids. De site wordt volledig gevuld door vrijwilligers die hiervoor geen enkele compensatie ontvangen. Het zwakste punt van DMOZ is de structuur: heel trots meldt de site zelf dat men over 590.000 categorieën beschikt. Dat is echter niet iets om blij mee te zijn. Het gebruik van onbezoldigde vrijwilligers heeft daarnaast nog een ander nadeel: het duurt soms geruime tijd voordat nieuwe sites worden toegevoegd aan de index. Ook fouten in de categorie-indeling worden niet altijd even prompt hersteld. Soms zwerven er op DMOZ zelfs rubrieken rond die allang niet meer in gebruik zijn.

### Google Directory – [directory.google.com](http://directory.google.com)

De Google Directory baseert zich op de inhoud van DMOZ, waaraan de eigen zoektechnologie is toegevoegd. Het voordeel daarvan is dat de informatie uit DMOZ ook via een invoerveld is te doorzoeken. Door de kracht van Google's zoektechnologie levert dit veel betere resultaten op dan bij andere internetgidsen die deze mogelijkheid bieden. Wie op 'architectuur' zoekt, zal bijvoorbeeld door Google worden geadviseerd naar de rubriek World/Nederlands/Kunst/Architectuur te gaan, waar allerlei koppelingen over dit onderwerp gesorteerd bij elkaar staan. Als Google een dergelijk advies kan geven, staat dat bovenaan het venster met zoekresultaten onder de noemer Gerelateerde categorieën. Het nadeel daarvan is dat Google een onhebbelijkheid van DMOZ heeft overgenomen: de zoekfunctie verwijst u soms naar rubrieken die geen dienst meer doen.



**Afbeelding 5.2** Onderaan de pagina kunt u ook de Nederlandse versie van de Google Directory kiezen. Rechtstreeks naar [directory.google.nl](http://directory.google.nl) gaan, levert helaas hetzelfde resultaat op als [directory.google.com](http://directory.google.com).



#### Meer sites

Door bij Google Directory de naam van een u al bekende website in te vullen, kunt u vaak eenvoudig meer sites over hetzelfde onderwerp vinden.

### LookSmart – [www.looksmart.com](http://www.looksmart.com)

Een goede internetgids die de nadruk legt op de mogelijkheid om via een invoerveld te zoeken. Net als Google Directory komt LookSmart vervolgens met suggesties van mogelijk interessante categorieën. Die worden door LookSmart wel enigszins verstoep: ze staan onderaan de pagina bij Related Directory Categories. Zoeken via een rubrieksindeling kan ook. Onder het invoerveld staan de twaalf categorienamen waarmee zo'n zoektocht bij LookSmart begint.

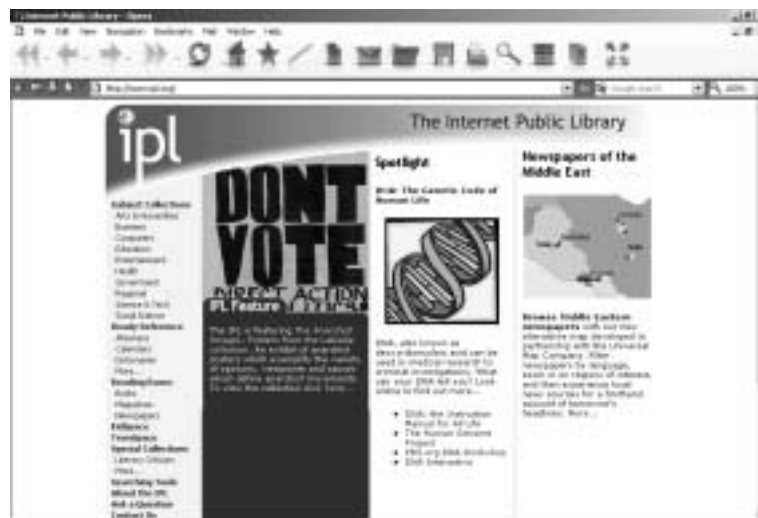


**Afbeelding 5.3** Onder de knop Articles, rechts bovenaan, bevindt zich een groot tijdschriftenarchief.

## Hoofdstuk 5 – Internetgidsen en het diepe web

### Internet Public Library – [www.ipl.org](http://www.ipl.org)

Deze site bevat een zeer kritische internetgids van de University of Michigan School of Information. De gids wordt samengesteld door professionals en dat is te merken. De makers van de Internet Public Library (IPL) streven overduidelijk niet naar kwantiteit, maar naar kwaliteit. Dat is tegelijkertijd het zwakke punt van IPL, waar soms essentiële sites ontbreken die door de samenstellers wellicht te gewoontjes werden geacht. IPL is niettemin een uitstekend startpunt voor met name meer academisch getinte zoektochten. Leuk zijn de specials, vaak geënt op een actueel onderwerp.



**Afbeelding 5.4** De specials leiden soms af van de internetgids in de menu-balk aan de linkerkant van het scherm.

### Library's Index to the Internet – [www.lii.org](http://www.lii.org)

De naam zegt het al: deze internetgids is gemaakt door en voor bibliothecarissen. Met de kwaliteit en de structuur van de site zit het dus wel goed en als er dan volgens oud-Hollandse traditie gezeurd moet worden, dan kan hoogstens worden gezegd dat de makers van de Library's Index to the Internet (LII) wel erg fanatiek bezig zijn. Er worden namelijk met grote regelmaat nieuwe interessante sites aan LII toegevoegd. Interessant is dat de site ook voorziet in een nieuwsdienst, waardoor geïnteresseerden automatisch op de hoogte kunnen blijven van nieuw toegevoegde sites. Dat kan via een traditionele nieuwsbrief per e-mail of per zogeheten RSS-feed (zie hoofdstuk 10 over het gebruik van RSS).



**Afbeelding 5.5** Door rechts bovenin het venster *Subscribe* te kiezen, kunt u toevoegingen aan de LII per e-mail toegezonden krijgen.

**Library of Congress – [international.loc.gov/www.loc.gov](http://international.loc.gov/www.loc.gov)**

Hoe u ook over de Amerikanen denkt, deze site van de Amerikaanse volksvertegenwoordiging mag u niet missen. Professionele documentalisten hebben een werkelijk uitstekende internetgids gemaakt. Het meest interessant is de internationale internetgids op [international.loc.gov](http://international.loc.gov). In de sectie *Portals to*



**Afbeelding 5.6** Behalve *Portals to the World* is ook de sectie *Research guides & databases* aan te raden.

## Hoofdstuk 5 – Internetgidsen en het diepe web

the world treft u hier mini-internetgidsen aan over vrijwel alle landen ter wereld. Op de hoofdpagina van de Library of Congress ([www.loc.gov](http://www.loc.gov)) is het gedeelte Index A-Z aan te raden. Hier houdt zich ook veel interessante informatie schuil, maar deze extraheren vereist wat meer werk.

### Resource Discovery Network – [www.rdn.ac.uk](http://www.rdn.ac.uk)

Een internetgids die zich richt op academische informatie. Resource Discovery Network (RDN) verzamelt websites in een aantal wetenschappelijke disciplines, variërend van Bioengineering tot aan Ethnography and Anthropology. Ietwat vaag is de structuur van de site wel: soms krijgt u de keuze uit nogal onduidelijke categorieën als Primary sources en Secondary sources. Daar komt nog bij dat het RDN zoveel koppelingen bevat, dat u regelmatig zelfs na het navigeren door de diverse rubrieken uiteindelijk een overvloed aan sites op uw scherm krijgt. Het handigste zoekt u daarom in RDN door eerst een zo specifiek mogelijk vakgebied te kiezen (bijvoorbeeld English literature). Wanneer u daarmee niet meer verder komt, voert via het veld bovenin het venster een specifiek trefwoord in om verder te zoeken. Let op: begin niet al in het hoofdvenster met zoeken via het invulveld, want dan zoekt u in alle rubrieken tegelijkertijd.



**Afbeelding 5.7** Een zoektocht via RDN begint door een van de rubrieken bij Browse te kiezen.



## Het diepe web

Het diepe web wordt ook wel het onzichtbare web genoemd, en dat heeft een goede reden. Het gaat hier om dat gedeelte van internet dat onzichtbaar is voor zoekmachines. Inderdaad, er zijn delen van internet die u wel kunt zien, maar die voor een zoekmachine verborgen blijven. Dat is niet altijd zo geweest: in vroeger tijden was veel meer van internet 'zichtbaar' – al waren zoekmachines in die tijd nog niet zo goed om daar optimaal van te kunnen gebruikmaken. Het onzichtbare web wordt steeds groter: volgens een schatting van de Amerikaanse internetpublicatie Salon.com uit maart 2004 is inmiddels al 99 procent van het web niet meer zichtbaar voor zoekmachines.

In één regel samengevat bestaat het onzichtbare of diepe web uit informatie die wel *via* websites te benaderen is, maar niet echt *op* die websites staat, maar bijvoorbeeld op computersystemen die daarachter hangen.

Om te begrijpen hoe dat kan, is het handig om het internet van weleer te vergelijken met een wildwest-stadje aan het einde van de negentiende eeuw, bestaande uit één hoofdstraat waarlangs alle belangrijkste winkels zich bevonden. Wie een inventarisatie wilde maken van de goederen die de diverse winkels aanboden, kon volstaan met een wandeling door de hoofdstraat en een grondige blik in de diverse etalages. De persoon die door de hoofdstraat loopt met een notitieblokje om deze inventarisatie te maken, is in dit voorbeeld uiteraard de zoekmachine; de winkels en hun voorraad staan respectievelijk voor websites en de aangeboden informatie.

Inmiddels is dat wildwest-stadje veranderd in een metropool à la New York. Er zijn meer winkels dan ooit tevoren, maar hun voorraad is niet zo eenvoudig meer te inventariseren. Wat er in de etalages staat, is maar een heel klein deel van het werkelijke assortiment. Het grootste deel staat in het magazijn. Weliswaar kan het bedienend personeel van de winkels zonder problemen de gewenste goederen leveren, maar alleen als een klant naar binnen stapt en erom vraagt. Het spreekt voor zich dat het personeel nogal geïrriteerd zou raken als degene die de inventarisatie maakt (de zoekmachine) de vraag zou stellen om de gehele magazijnvoorraad maar even op te dreunen.

Zo gaat het ook op internet. Vroeger waren websites simpel: ze bestonden uit een aantal pagina's, vaak geschreven in de internetopmaaktaal HTML. Dit waren zogeheten statische pagina's, die niet of nauwelijks aan verandering onderhevig waren. Een zoekmachine kon daar relatief eenvoudig greep op krijgen. Tegenwoordig bestaan heel veel sites niet meer louter uit dergelijke statische pagina's. In plaats daarvan is het statische gedeelte beperkt en louter nog bedoeld om toegang te geven tot achterliggende informatie. Dit is

## Hoofdstuk 5 – Internetgidsen en het diepe web

meestal een database oftewel een kaartenbak met gegevens. De bezoeker van de site gebruikt vaak een invoerveld om op te geven welke informatie deze wil, waarna hij of zij de gegevens te zien krijgt. Deze pagina's zijn dynamisch. Dat wil zeggen dat ze louter op uw verzoek zijn gemaakt en ook niet langer bestaan dan de tijd waarin u ze weergeeft. Een zoekmachine zal ze nooit zien en heeft zelf ook geen toegang tot een achterliggende database. De enige manier waarop een zoekmachine te weten zou kunnen komen welke informatie de achterliggende database van een website bevat, is de website te gaan 'bevragen' op alle mogelijke letter- en woordcombinaties.

Een voorbeeld. Als u op Google zoekt naar Kamerhandelingen "Delft Instruments" levert dat exact nul treffers op. Gaat u echter naar de website Officiële Publicaties van de Rijksoverheid ([www.overheid.nl/op](http://www.overheid.nl/op)) en zoekt u daar op "Delft Instruments" (met alle zoekopties aangevinkt), dan treft u legio documenten aan over het Delftse bedrijf dat ooit in het nieuws kwam als leverancier van nachtkijkers aan het Irak van Saddam Hussein.



Afbeelding 5.8 Voor al uw kamerstukken.

Nog een voorbeeld. Stel, u zoekt wetenschappelijke studies over griep in Nederland, oftewel influenza. Als u bij Google *influenza netherlands scientific studies* typt, krijgt u weliswaar veel kaf te zien, maar is de oogst aan relevante resultaten mager. Gaat u echter naar PubMed ([www.pubmed.gov](http://www.pubmed.gov)) en typt u daar *influenza netherlands*, dan hebt u gegarandeerd meer succes. Op deze site staan namelijk ruim 14 miljoen medisch-wetenschappelijke studies bij elkaar. Die zijn niet allemaal gratis te bekijken, maar de samenvattingen meestal wel. Google kent de meeste



**Afbeelding 5.9** Veel gepubliceerde medisch-wetenschappelijke artikelen komen in PubMed terecht, al zijn ze niet allemaal gratis te bekijken.

studies in PubMed niet. Hoogstens wanneer een arts zo ijdel is geweest zijn in PubMed opgenomen studie ook nog op zijn eigen website op te nemen, hebt u met Google nog wel eens een kansje.

De vraag luidt nu: hoe kunt u toch zoeken in het diepe web. Het antwoord, ietwat teleurstellend, is: uiterst moeizaam.

Uit de voorgaande twee voorbeelden is al wel duidelijk geworden dat u om in het diepe web te zoeken, moet weten welke sites toegang geven tot bepaalde gedeelten van dat diepe web. Zo geeft PubMed toegang tot een medisch-wetenschappelijk gedeelte en geeft de website Officiële Publicaties toegang tot stukken van onder meer het Nederlandse parlement.

In welke situatie u via het diepe web moet zoeken en hoe u dat het beste aanpakt, kunt u bepalen met de volgende drietrapsstrategie.

- 1 Zoek specifiek.** U begint uw zoektocht zoals u in hoofdstuk 2 hebt geleerd: door specifiek te zoeken via Google. Wellicht kunt u zich zo veel tijd besparen alvorens u op het diepe web te richten.
- 2 Zoek breed.** Google heeft niets opgeleverd, andere relevante zoekmachines ook niet. Nu moet u een diepe-website proberen te vinden die wel de informatie bevat die u zoekt. Het kenmerkende van diepe-websites is als gezegd dat de meeste informatie niet zichtbaar is voor zoekmachines. Slechts de 'voorgevel' van een dergelijke site is te vinden via

## Hoofdstuk 5 – Internetgidsen en het diepe web

zoekmachines. Op deze pagina's staat vaak elementaire informatie over de inhoud van de site die wel in zoekmachines wordt opgenomen. Daarvan kunt u dan ook gebruikmaken. Anders gezegd: u kunt de website Officiële Publicaties niet vinden door op Delft Instruments te zoeken. Maar als u bij Google "Officiële Publicaties" overheid had getypt, was u wel op de site Officiële Publicaties geëindigd. Een ander voorbeeld: als u op een zeer specifieke psychiatrische aandoening zoekt via Google, maar nul op het rekest krijgt, probeer dan een algemenere zoekvraag als "mental health" om zo te proberen een overzichtssite over een bepaald onderwerp te vinden

- 3 Zoek nogmaals specifiek.** U hebt de betreffende site gevonden die toegang geeft tot het stukje diepe web waar zoekmachines niet bij kunnen. Hier dient u uw oorspronkelijke specifieke zoekvraag opnieuw te stellen. Levert dat niets op, ga dan terug naar de vorige stap en probeer meer diepe-websites te vinden over hetzelfde onderwerp.



### Maak een bladwijzer

Het zoeken naar diepe-websites kan een langdurig en soms frustrerend proces zijn. Daarom is het aan te raden om onmiddellijk een bladwijzer te maken van nuttige diepe-websites over onderwerpen waarnaar u vaak research doet.

De voorgaande strategie is bedoeld als een algemene methode om diepe-websites te vinden. Er zijn echter ook in het diepe web gespecialiseerde sites die u hierbij een handje kunnen helpen. Let op: voor alle sites die hier besproken worden, geldt uiteraard dat u begint met zoeken vanaf stap 2: Zoek breed.

### Infomine – [infomine.ucr.edu](http://infomine.ucr.edu)

Een werkelijk uitstekende site met naar eigen zeggen een krappe 120.000 academische hulpbronnen op allerlei gebied. Deze site kan u helpen met het vinden van diepe-websites op allerlei gebieden. Om het meeste rendement te halen uit Infomine doet u er goed aan eerst een rubriek te selecteren en daarna het invoerveld te gebruiken om te zoeken. Gebruik stap 2 uit de diepe-webzoekstrategie: zoek dus op "European Union" 1aw als u op zoek bent naar een Europese richtlijn, en niet naar Directive 95/46/EC.

### The Invisible Web – [www.invisible-web.net](http://www.invisible-web.net)

Deze site is in feite een gratis voorproefje van het gelijknamige boek, dat vanwege zijn sterke Amerikaanse inslag minder geschikt is voor de Nederlandse markt. Niettemin is de site interessant, en dan vooral omdat u er niet



**Afbeelding 5.10** Selecteer voor de beste resultaten eerst een van de rubrieken voordat u een zoekwoord typt.



**Afbeelding 5.11** De site van *The Invisible Web* bespaart u de kosten van het boek.

voor hoeft te betalen. Het beste brengt u een eenmalig bezoek aan deze site en besteedt u vervolgens een uurtje aan het doorpluizen van de rubrieken die mogelijk voor u interessant kunnen zijn. Maak vervolgens bladwijzers van de sites die relevant zijn. In een later stadium hebt u hier profijt van.

## Hoofdstuk 5 – Internetgidsen en het diepe web

### Complete Planet – [www.completeplanet.com](http://www.completeplanet.com)

Noemt zichzelf de Deep Web Directory, maar dat is te veel eer. Complete Planet had tot voor kort een slechte reputatie omdat het de diepe-websites wat al te gretig vermengde met sites van adverteerders. Inmiddels is dat enigszins verbeterd, maar Infomine geniet nog altijd de voorkeur. Het zoek-algoritme waarmee Complete Planet de eigen kaartenbak met diepe-websites doorploegt is namelijk verre van perfect. Niettemin is deze site een goede tweede om te gebruiken nadat de mogelijkheden van Infomine zijn uitgeput.



**Afbeelding 5.12** Complete Planet, aan de beterende hand na al te veel commerciële vervuiling.

### Beaucoup – [www.beaucoup.com](http://www.beaucoup.com)

In het vorige hoofdstuk kwam deze site al aan de orde als overzichtssite van zoekmachines. Als u zoekt via het invoerveld, door Beaucoup weinig bescheiden Beaucoup Super Search genoemd, komt u echter ook op diepe-websites terecht. Zoals in hoofdstuk 4 al is gemeld, moet u wel oppassen voor de commerciële uitspattingen van Beaucoup, dat ooit wel eens heeft gehoord van redactionele onafhankelijkheid, maar daar niets mee te maken wil hebben.