

Het zoeken naar een speld in een hooiberg

Het aanbod aan informatie op Internet is overweldigend, maar net dat te vinden waar men naar op zoek is, blijkt een hele kunst. Het inschakelen van één van de bekende zoekmachines (bijv. Alta Vista, Yahoo) lijkt een goede oplossing, maar de resultaten vallen vaak tegen omdat de opbrengst daarvan niet specifiek genoeg is of er soms zelfs helemaal naast zit. In deze bijdrage zullen een aantal effectievere methoden worden beschreven en zal worden gewezen op een Nederlandse tandheelkundige 'verzamel'-website, een zogenaamde startpagina.

Metazoekmachines

Wie op Internet regelmatig een zoekmachine gebruikt, zal hebben gemerkt dat het invoeren van een zelfde trefwoord verschillende resultaten kan opleveren. Dat heeft te maken met de gebruikte zoektechniek en de database waarin gezocht wordt. Het is alsof men met een grofmazig sleepnet de zee bevest: alleen de grote vissen dicht bij de oppervlakte worden gevangen. Door de mazen kleiner te maken, met andere woorden trefwoorden te combineren met Boleaanse operatoren, wordt het resultaat beter, maar wordt toch nog veel van de dieperliggende informatie gemist. Door het inzetten van meerdere zoekmachines tegelijkertijd wordt het bereik vergroot. Deze zogenaamde metazoekmachines produceren zo snel de eerste tien à vijftien relevante zoekresultaten van iedere machine, dat het lijkt alsof men slechts één machine aan het werk heeft gezet. Een voorbeeld hiervan is MetaCrawler.

Een aparte categorie vormen de 'directories', zoals Yahoo. Hier kan op twee manieren worden gezocht: door het invoeren van een zoekterm, maar ook via een boomstructuur. Elke categorie is verdeeld in subcategorieën, die op hun beurt weer onderverdeeld zijn tot men uiteindelijk op de gezochte pagina komt. Zo'n 'directoriestructuur' wordt niet door software aangemaakt maar door een redactie, wat subjectief en vertragend werkt.

Installeren van eigen metazoekmachine

Tegenwoordig is het mogelijk op de eigen computer software voor het zoeken te installeren. Een voorbeeld hiervan is het programma 'Copernic 2000'. Dit programma combineert tien van de bekendste zoekmachines en is voorzien van een aantal voorgedefinieerde rubrieken. Een freeware versie is te downloaden op www.copernic.com. Tegen geringe kosten zijn tevens een Plus- en een Pro-versie verkrijgbaar, waarin men zelf zoekmachines kan toevoegen. Hiermee is het mogelijk een groot aantal van deze programma's samen aan het werk te zetten. Copernic is erg gebruikersvriendelijk en levert goede en snelle resultaten. Bij dit alles dient men in het oog te houden dat de traditionele zoekmachines indexeren op html-webpagina's. Deze webpagina's moeten het vooral van de hypertextlinks hebben om 'gevonden' te kunnen worden. Tegenwoordig wordt veel informatie in databases opgeslagen, die via een opvraagtaal binnen de website worden benaderd. Als dan geen hypertextlink met een beginpagina (bijv. homepage) aanwezig is, blijft deze verborgen.



Openings scherm website CompletePlanet.

DeepWeb

Onderzoek in Amerika heeft aangetoond dat zoekmachines slechts éénvijfhonderdste deel van de Internetpagina's indexeren. Het gaat naar schatting om een miljard pagina's. De circa 500 miljard overige pagina's met vaak goede informatie bevinden zich diep weggestopt in databases die niet worden gevonden. Dit zogenaamde 'DeepWeb' is onderwerp van onderzoek door BrightPlanet, waarvan de website is te bezoeken met het adres www.completeplanet.com. Op deze website vindt men de verantwoording van de genoemde cijfers en wordt beschreven met welke methoden de verborgen informatie mogelijk aan de oppervlakte kan worden gebracht. Hoewel deze nieuwe zoekmethode, 'LexiBot' geheten, nog in een ontwikkelingsfase (zgn. betafase) verkeert, zijn op de site reeds een aantal DeepWebsites toegankelijk. Sommige van de gevonden databases vereisen echter een toegangscode die of registratie en/of kosten met zich meebrengt.

Interessant is nog om te vermelden dat volgens de onderzoekers van BrightPlanet zelfs met de snelste metazoekmachines tegenwoordig niet meer dan 45% van de html-pagina's wordt gevonden. Bovendien ontstaat door de buitengewoon snelle aanwas van informatie een groeiende achterstand in indexeren, waardoor dit percentage elk jaar afneemt.

Websites met startpagina's

Wie niet precies weet waar en hoe te zoeken, kan gebruikmaken van zogenaamde startpagina's. Dat zijn websites die per categorie gerangschikte sites verzamelen en in een menu van onderwerpen, startpagina's genaamd, aanbieden. Een Nederlands voorbeeld is www.startpagina.nl. De categorie tandheelkunde is rechtstreeks te vinden met <http://tandarts.pagina.nl>. Bij het openen van de tandartspagina krijgt men achttien rubrieken te zien, die samen circa tweehonderd links naar tandheelkundige onderwerpen bevatten. Men kan zelf een site aanmelden en onder een gekozen rubriek laten opnemen. De kwaliteit van de sites varieert dan ook van reclame tot echte informatiebron. Voor een ieder die snel onder de kop 'tandheelkunde' wenst te surfen, kan op deze startpagina beginnen. Een benadering die op een donkere winteravond best onderhoudend kan zijn. Wie echter werkelijk op zoek is naar bepaalde informatie ontkomt er niet aan zich te bekwaamen in de zoektechnieken die voor de verschillende zoekmachines nodig zijn. De energie en de tijd die men hiervoor investeert, wordt vroeg of laat beloond met een efficiëntere wijze van informatie vergaren op het Internet.

S.L. Liem, redacteur Internet